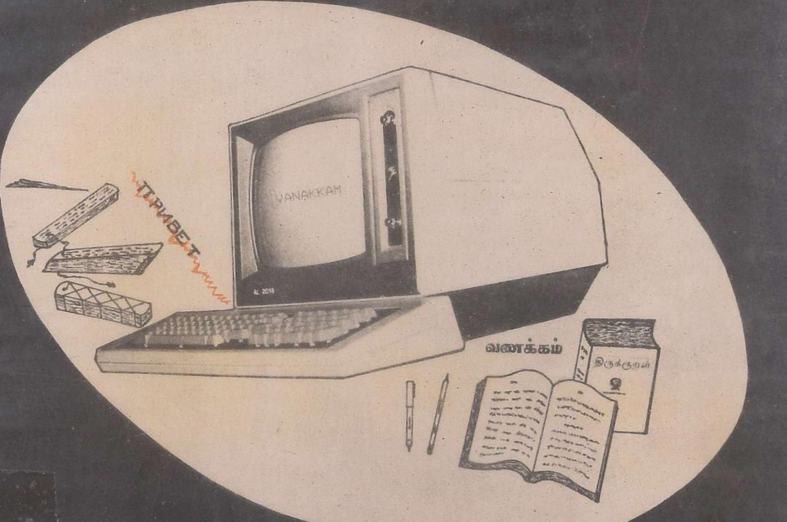
RESEARCH REPORT
(PHASE — I)

TAMIL UNIVERSITY MACHINE TRANSLATION SYSTEM (TUMTS)

RUSSIAN — TAMIL



0818
2D04
818

K.C.Chellamuthu
K.Rangan
K.Murugesan

Tamil University
Thanjavur

TAMIL UNIVERSITY MACHINE TRANSLATION SYSTEM (TUMTS)

MAHARAJA SERFOJI'S SARASWATI MAHAL LIBRARY
THANJAVUR

RUSSIAN - TAMIL

RESEARCH REPORT — PHASE-I

K.C.CHELLAMUTHU
K.RANGAN
K.MURUGESAN

TAMIL UNIVERSITY
THANJAVUR
S.INDIA

Tamil University Machine Translation
System (TUMTS)
Russian — Tamil

First Edition November, 1984
Tamil University
Thanjavur

D6,8(B)
M4

Tamil University offset Press
Thanjavur, Tamil Nadu
India

PREFACE

The fast development in the areas of science and technology has created a new situation in which the progress of human civilization crucially depends on the dissemination of information. The development of human society is closely linked with the development of its language also. Information must be accessible to the members of a society in the language they use. Here comes the significant role to be played by translation. When translation involves large quantity of materials at low cost, the computer is brought to the scene. A brief history of Machine Translation given on page (1), presents an outline of the attempts initiated in various countries and also the fall and rise of the Machine Translation.

At the suggestion of our Vice-Chancellor prof. V.I.Subramoniam, a pilot project on "Computer-aided-Translation from Russian into Tamil" was initiated in October 1983. A small team was formed drawing the scholars from the Departments of Translation and Linguistics. The team had an expert in Russian language (K.Murugesan, Department of Translation), a computer engineer (K.C. Chellamuthu, Data centre) and a linguist (K. Rangan, Department of Linguistics). It was decided to undertake this

study as an experimental one to experience the problems directly. We believe that the pilot work undertaken could help us to conceive the project in a bigger way at later stage. The work progressed with the guidance and supervision of our Vice-Chancellor Prof. V.I. Subramoniam.

The Data centre has a Microprocessor based Fourth Generation Computer with a working memory of 64 K.bytes and a dual floppy drives of one million bytes of auxiliary storage. Besides this, the hardware consists of a Dot matrix printer with a speed of 150 cps. The software system includes a number of language processors with powerful software modules. This machine was available to the team to work on this project.

Word order is relatively free in both Tamil and Russian. The markers indicating different case relations occur after the nouns in both the languages. It is believed that the transfer rules which convert the surface syntactic structure of the source language into the surface syntactic structure of the target language may be less when Russian is selected instead of English.The attempt is Russian --->Tamil.

THE TAMIL NADU GOVT. MAHARAJA SERFOJI'S SARASWATHI MAHAL LIBRARY THANJ.VUR

There were certain constraints under which the team was to carry out the research work. First and foremost was the constraint on the power of the memory of the machine. Storing the lexical items of Russian and Tamil and the elaborate grammatical rules of these two languages became a crucial problem for the team to work. Soon it was realized that the memory power of the machine had to be increased to store text oriented/fieldwise dictionary. It will be serious when we take more texts in various domains for the purpose of translation.

The present work is inadequate and incomplete in many respects. First is the coverage of data. The present system does not include the sentences with future tense. Similarly complex sentences such as co-ordination, conditional etc., are not included in this phase of translation work. Inclusion of more texts from different domains may require further revision of the present system. The report presented here is incomplete and tentative. Since the upper case letters are used for transcribing the language materials of Tamil and Russian, we find the scheme of transliteration defective in certain respects. There is no way to

distinguish ல ,ள and ழ in our scheme. The same is the case with ட and த , ர and ற and ன,ண,ஞ் and ங் using both the upper and lower case letters, the scheme of transliteration will be revised in the second phase of the report. It is prepared with a view of sharing our experiences with fellow scientists working in the area of Computer-aided-translation.

We believe sincerely that the suggestions to be received will not only help us to realize the inadequacies of the present work but will also contribute to envisage and generate larger projects to be undertaken in the area of Computer-aided-Translation. Other Indian languages may also be included in this project at later stages.

We think that it is appropriate to recall the scholars and other non-academic staff of the Tamil University who helped us in many ways. First we remember the days of our discussion with our Vice-Chancellor Prof. V.I.Subramoniam who showed keen interest in the progress of this research project. We record our debt of gratitude for his patient guidance and supervision of this project at every stage. The staff of the Data centre

extended their help in all possible ways. To them we like to express our thanks. Finally we express our thanks to Mr.J.A.Dias for the neat execution of typing work.

THANJAVUR-1
25-11-1984 .

K.C.CHELLAMUTHU
K.RANGAN
K.MURUGESAN.

CONTENTS

	Preface
	Abbreviations
	System of Transliteration
1	Linguistic Aspects of Automatized Translation
1.1	A brief history of Machine Translation
1.2	Scope and Aim
1.3	Components
1.4	Search for equivalents
1.5	Comparison of Russian and Tamil
1.6	Word Order
1.7	Area of Problems
1.8	Transfer rules
2	An Outline of Russian and Tamil Structure devised for Machine Translation
2.1	Procedure
2.2	Formation of dictionary
2.3	Morphology
2.3.1	Nouns
2.3.1.1	Singular versus plural

2.3.1.2 Cases

2.3.2 Adjectives

2.3.3 Verbs

2.3.3.1 Present tense

2.3.3.2 Past tense

2.3.3.3 Future tense

2.3.4 Adverb

2.4 Syntax

2.4.1 Sentence pattern

2.4.1.1 Active Versus Passive

2.4.1.2 Passive Sentences with short participles and adjectives.

2.4.1.3 A Russian Passive sentence with a short form of adjective

2.4.2 Degrees of Comparison

2.5 Automatic synthesis of text in a natural language

2.6 Complex sentences

3 Programming Aspects of Machine Translation system

3.1 Introduction

3.2 Computing resource

3.3	System modelling
3.4	Pre-processor system
3.4.1	Input text entry
3.4.2	Source Text Editing Module
3.5	Pre-processor Module
3.5.1	Scanner
3.5.2	Parser
3.6	Language Data base
3.6.1	Structure of the Data bases
3.7	Translation phase
3.7.1	Verb Phrase Analyser
3.7.2	Non Verb Phrase Analyser
3.7.3	Syntax Analyser
3.7.4	Lexical Analyser
3.7.5	Morphological Analyser
3.7.6	Resolving Semantic ambiquities
3.7.7	Word to word Translation
3.8	Generation Phase
3.8.1	Generation Module
3.8.2	Structural Transfer Module

3.9 Post editing phase

3.10 Output Phase

3.11 Case study

3.11.1 Experimental Observation

3.11.2 Programming modules and files

3.12 Conclusion

APPENDIX

Appendix-A		Block diagram of TUMTS System
"	B	Bilingual Dictionary
"	C	Types of declensions
"	D	Original Russian Text
"	E	Input Text
"	F	Pre-edited Text
"	G	Translated Tamil text (Romanised)
"	H	Translated Text in Tamil.

Bibliography.

ABBREVIATIONS

acc	=	accusative
ADJ	=	Adjective
CP	=	Comparative phrase
Com	=	comparative
FEM	=	Feminine
GEN	=	Genitive
H	=	Head
Hon	=	Honorific
$L_{1/2}$	=	Language
LOC	=	Locative
MAS	=	Masculine
N	=	Noun
NEUT	=	Neuter
NP	=	Noun phrase
O	=	Object
PL	=	Plural
PROX	=	Proximate
REM	=	Remote

R	=	Russian
S	=	Subject
SING	=	Singular
SOC	=	Sociative
T	=	Tamil
V	=	Verb
VP	=	Verb Phrase
MT	=	Machine Translation
HT	=	Human Translation
MAHT	=	Machine Aided Human Translation
HAMT	=	Human Aided Machine Translation
SYSTRAN	=	System Translation
CULT	=	Chinese University Language Translator
GETA	=	Groupe D'Etudes pour La Traduction Automatique
TUMTS	=	Tamil University Machine Translation System.

SYSTEM OF TRANSLITERATION

TAMIL

அ	=	A	ஞ்	=	N
ஆ	=	AA	ட்	=	T
இ	=	I	ண்	=	N
ஈ	=	II	த்	=	T
உ	=	U	ந்	=	N
ஊ	=	UU	ப்	=	P
எ	=	E	ம்	=	M
ஏ	=	EE	ய்	=	Y
ஐ	=	AI	ர்	=	R
ஒ	=	O	ல்	=	L
ஓ	=	OO	வ்	=	V
ஔ	=	AU	ழ்	=	L
க்	=	K	ள்	=	L
ங்	=	N	ற்	=	R
ச்	=	C	ன்	=	N

RUSSIAN

Аа	=	A	Пп	=	P
Бб	=	B	Рр	=	R
Вв	=	V	Сс	=	S
Гг	=	G	Тт	=	T
Дд	=	D	Уу	=	U
Ее	=	E	Фф	=	F
Ёё	=	JO	Хх	=	H
Жж	=	ZH	Цц	=	C
Зз	=	Z	Чч	=	CH
Ии	=	I	Шш	=	SH
Йй	=	J	Щщ	=	SHCH
Кк	=	K	Ъъ	=	:
Лл	=	L	Ыы	=	Y
Мм	=	M	Ьь	=	J
Нн	=	N	Ээ	=	E
Оо	=	O	Юю	=	JU
			Яя	=	JA

CHAPTER-1

LINGUISTIC ASPECTS OF AUTOMATIZED TRANSLATION

Automatized Translation or familiarly known as Machine Translation or Computer-aided-Translation is an interdisciplinary activity which involves the expertise from the fields of language, linguistics and computer programming. The departments of Translation and Linguistics initiated an experimental study on 'Russian-Tamil Machine Translation'. The study was conceived as a pilot one and therefore the objective was kept as modest as possible.

Since the researchers are aware of the rise and fall of the Machine Translation projects in various parts of the world, the objective of the study is practical. It was designed as an experimental work to experience/face the problems directly. The experience gained in the pilot study will be helpful to conceive and propose a long term project in this area.

1.1 A BRIEF HISTORY OF MACHINE TRANSLATION (MT)

Even in 50's, an attempt was made to use the computer for the purpose of translation. The memorandum issued by Warren Weaver was the initial

point and it was followed by an attempt of Y.Bar-Hillel at the Massachusetts Institute of Techno - logy. A longer project was undertaken by the University of Georgetown. It was a project of Russian to English Machine Translation(MT) system. In 1954, both the University of Georgetown and IBM convinced the authorities to fund on a large scale for the projects of MT. At Moscow, a demonstra - tion on French to Russian programme was given by O.S. Kulagina and I.A. Mel'chuk in 1956.

One may discern two kinds of trends in the research activities of MT. First, there was an increasing demand for translation. Information acquirement was the basis for the requirement of large amount of translation. The main goal to look for the machine was to get the translation in many restricted topics at the lowest cost possible. Secondly it was rather viewed as a source for deep research.

It is found that there are two kinds of tendencies in MT. First is a word for word substitution with adhoc routines and rearrangement of the translated sentences in the target language.In the second, the translation is effected through an interpreted structure of the input text.

At the requests of physicists, Russian to English translation of scientific literature was attempted by the scholars of first generation system. P.Toma developed a commercial system called as SYSTRAN (system translation latsec, Inc., California) after several years of research. It had a great impact on the market. The major advantage of a Machine Translation system is its speed. According to the statistics available the SYSTRAN could be able to translate 100,000 words in about 5 minutes and during 1974 alone the system has translated about 24,000 pages of texts for US air force. The second generation system addressed to the question-answer systems and man-machine communication by means of natural languages.

In 1971, Y.Bar-Hillel presented a report in which he argued for the goal to be bounded by reasonable limits. The report was contributed to the seminar at the University of Texas. The LOGOS development corporation in USA developed an English to Vietnamese translation system called LOGOS in 1973. A system called CULT (Chinese University Language Translator) was developed by the Hong Kong University, which automatically translated Chinese newspapers into English. French Textiles Institute in Paris has developed an equally

important system called TITUS which translated simplified sentences between the four languages German, Spanish, French and English. At the University of Montreal, the TAUM project succeeded in producing the English to French weather report translation in 1975. A centre for automatized translation from English into Russian of documentation in electro technics was established in USSR in 1972-73. In 1976-77, the USSR funded the Central Office for Translation to initiate multi-lingual automatized translation systems.

According to the report received from Prof. B.Vauquois of the Universite Scientifique Et Medicale de Grenoble, Cedex a Machine Translation system called GETA (GROUPE D'ETUDES POUR LA TRADUCTION AUTOMATIQUE) was developed for Russian to French Translation. It is a joint venture of scientists from USA, Canada, Japan and France. GETA system is operational at the University Sains Malaysia at Penang, University chulalongkorn at Bangkok, computing centre of Academy of sciences at Peking.

In the first stage, the "information acquirement" played significant role in the demand of Machine Translation. When this is extended, a

newer aspect *"information diffusion"* plays important role in the activity of automatized translation. Diffusion of information is crucial in the areas of research, commerce, etc. It increases the demand for translation and therefore the translation is leading to third generation system. At the same time, it is to be emphasized that the goal set up for automatized translation must be reasonable and practical.

It is really interesting to note that, there has been a forward movement in the direction of Machine Translation of Indian Languages . It is imperative that a very favourable condition is created to carry out further investigations to develop Machine Translation systems for all the Indian Languages. Keeping this in view the Tamil University has initiated the Russian to Tamil translation system.

1.2 SCOPE AND AIM :

Language is a collection of dialects representing both the regional and social divisions of a speech community. Different styles used in the writings of literary and science texts also form part of a language. The main task before

the team was to select the style of materials, which could be fed into the computer. In other words, the nature of the input materials was to be selected. It was decided to exclude the spoken dialects from this study. The written materials were taken into consideration for the purpose of Machine Translation. The style that is used in the writings of literary discourses is considered to be complex and subjective. The term '*subjective*' is here used to refer to the expression that expresses the feelings and emotions.The style exhibits characteristic features of subjectivity.

As a consequence of the subjectivity of the content the sentence used are rather complex. In literary works, the aim is not just to convey the ideas or thoughts. The manner in which an idea or thought or feeling is expressed plays a crucial role in determining the complexity of a sentence at the level of performance.

Since science deals with external objects , the style preferred is impersonal. Language is here used for the purpose of communication in narrow sense. Expression of feelings or emotions does not find place in communication. Facts are

expressed objectively. Therefore the style is objective and not as complex as the style of literary discourses. The style used in the texts of science can be differentiated from the style used in the works of literature. One can easily discern the features associated with the style of science texts. So the team selected science texts from the Russian language as the input material for the computer. It was originally planned to select half a dozen text materials from various branches of science to improve the quality of translation. However it was decided to prepare a tentative preliminary report describing the problems faced, the results achieved and the methodology adopted for the purpose of computer aided translation with a single text. The report presented here is the first phase of the work and we believe that better results and further refinement in our methodology may be expected as we proceed to work with some more texts.

The following were the objectives :

a) To design a system for automatized Machine Translation of science and technical literature for a given pair of languages.

b) To compare the sentence structure in a pair of languages.

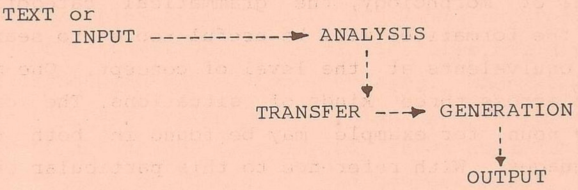
1.3 COMPONENTS:

From the point of view of linguistics, the activity of translation from L1 into L2 involves the search for equivalents. The search includes the areas of morphology, syntax and semantics. Dictionary part gives a list of lexical items with all necessary information in both the languages. It contains information not only with reference to the meaning of lexical items but also other information from other areas. One may visualise atleast three components in the process of translation. The text fed into the computer is first analysed. Analysis covers both the areas of morphology and syntax.

The structure of the sentences is represented in terms of syntactic labels such as NP, VP, V, N, ADJ, etc. Once the structure of the sentences of the text is analysed, it is then converted into the surface syntactic structure of the target language (L2).

The conversion from L1 into L2 is achieved by making use of the transformational/transfer rules. Transfer component, the second in the process of translation, consists a set of rules called transfer rules. These rules are based on a contrastive study of the surface syntactic structure of

L1 and L2. Generation is the third component of the translation process. It generates the actual sentences of the target language. This may be shown as

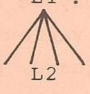


1.4 SEARCH FOR EQUIVALENTS :

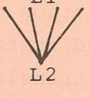
It was mentioned in the last section that the process of translation involves the search for equivalents in both the Source and Target languages. It may be explained by drawing the examples from Russian and Tamil. What is meant by equivalents in the process of translation ? With the help of linguistic theory we will attempt to explain this concept. The various levels postulated in the analysis of language provide convenient framework to search for the equivalents in both the languages. The levels postulated depends on the nature of a theory. The level of phonology is not taken into consideration in this work. The levels of morphology and syntax have been recognized for the purpose of translation work.

In an attempt to search for equivalents, it is important to recognize two things. One is conceptual and the other is form. For example both Russian and Tamil have nouns and verbs. From the level of morphology, the grammatical categories and the formatives will be useful tools to search for equivalents at the level of concept. One may come across three kinds of situations. The category noun, for example, may be found in both the languages. With reference to this particular category there is one-to-one correspondence. This may be shown

In the second kind of situation, one may come across with one-to-many correspondence. This may be shown as

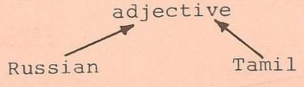


The third kind of situation will be many-to-one correspondence and this may be shown as

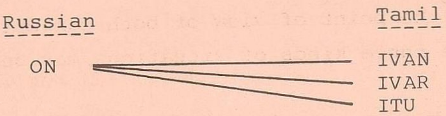


From the point of view of both concept and form, these three kinds of situations may be recognized.

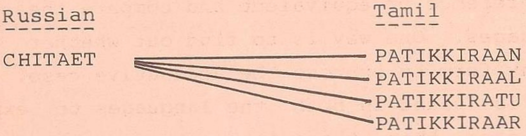
For the purpose of explanation, we will draw some examples from Russian and Tamil before we present systematically the similarities and differences between these two languages. When we examine the case system of Russian and Tamil, the concept accusative case may be taken as a point of reference or equivalent and compare these two languages. One way is to find out whether there is any form to express the accusative case. If a form is found in both the languages to express this case, then there is one-to-one correspondence at the conceptual level. Another example may be the presence of adjective as a grammatical category at the conceptual level. This may be shown as



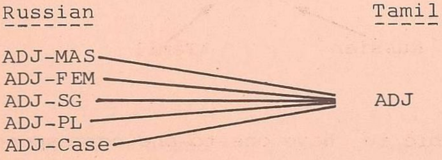
It is possible to have one-to-one correspondence at the conceptual level but many-to-one or one-to-many correspondence at the level of form. 'ON' is Russian is to be equated with 'IVAN', 'IVAR' and 'ITU'.



Similar situation may be found when we compare the person, gender and number marker of Tamil with Russian. AAN, AAL, TU and AAR are to be equated with only one form of Russian.



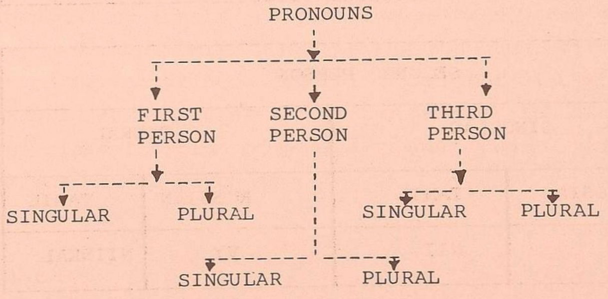
When we compare the forms of adjectives one finds many-to-one correspondence. The adjectives in Russian are inflected for the gender, number and case of the nouns.



It would become clear that the degree of success in translation is closely linked with a good contrastive analysis of these languages.

1.5 COMPARISON OF RUSSIAN AND TAMIL :

Before we took a science text for translation from Russian into Tamil, we selected some simple sentences from Russian and attempted to translate them into Tamil. As it has been pointed out earlier, either in terms of concepts (grammatical categories and grammatical formatives) or forms, the equivalents are established. Both in Russian and Tamil, we find three way distinction of personal pronouns. Number distinction (singular versus plural) is also maintained in the classification of pronouns. This may be shown as



From the diagram it is clear that the correspondence between Russian and Tamil in the

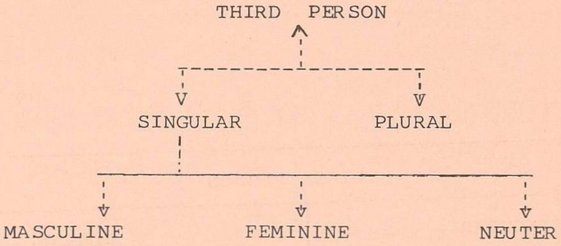
classification system of pronouns is one-to-one. It does not necessarily imply that the forms in these two languages are also in one-to-one correspondence. As far as the first person and second person are concerned, both these languages show similarity. It may be presented as follows :

FIRST PERSON			
SINGULAR		PLURAL	
RUSSIAN	TAMIL	RUSSIAN	TAMIL
JA	NAAN	MY	NAANKAL

SECOND PERSON			
SINGULAR		PLURAL	
RUSSIAN	TAMIL	RUSSIAN	TAMIL
TY	NII	VY	NIINKAL

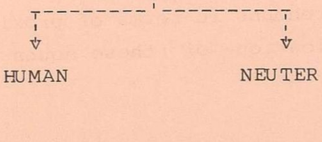
Both in Tamil and Russian singular versus plural distinction is maintained in the third person pronominal system. It may be shown as

RUSSIAN & TAMIL :



Under singular, masculine, feminine and neuter divisions are made in both Russian and Tamil. In Russian, no further distinction is made under plural. But in Tamil, human versus non human distinction is maintained. Plural is further distinguished as human plural and neuter plural in Tamil. The classification of plural nouns under third person may be given as :

TAMIL :

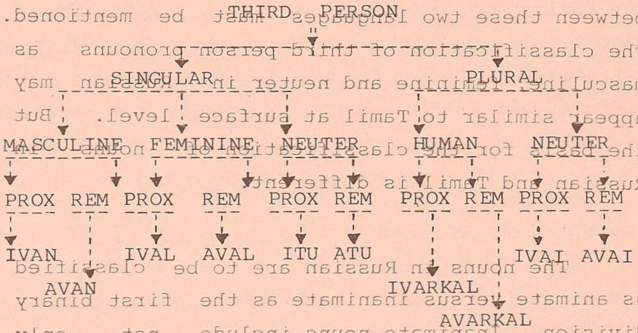


The following table presents the classification of pronouns of third person in both Russian and Tamil.

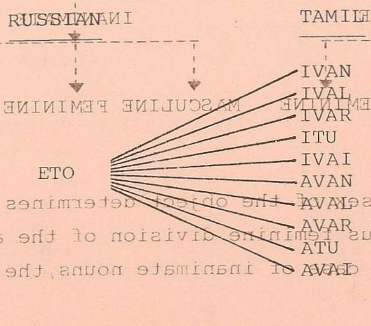
THIRD PERSON							
SINGULAR				PLURAL			
MASCULINE		FEMININE		NEUTER		RUSSIAN	TAMIL
RUSSIAN	TAMIL	RUSSIAN	TAMIL	RUSSIAN	TAMIL	--	HUMAN-NEUTER
ON	IVAN	ONA	IVAL	ONO	ITU	ONI	IVARKAL IVAI

The pronouns in Tamil may further be classified on the basis of remote and proximate. A-denotes remote and I-proximate. In third person one gets the following pronouns in terms of proximate and remote. Classification of these nouns may be shown as

TAMIL :

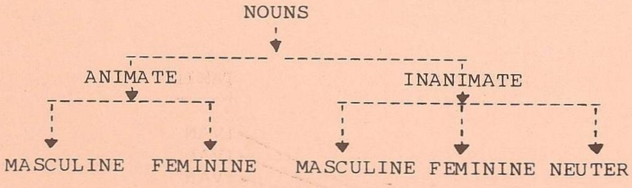


Surprisingly one comes across with the demonstrative pronouns ETO in Russian which is comparable to ten forms in Tamil.



Before we deal with other grammatical aspects of Russian and Tamil, one crucial difference between these two languages must be mentioned. The classification of third person pronouns as masculine, feminine and neuter in Russian may appear similar to Tamil at surface level. But the basis for the classification of nouns in Russian and Tamil is different.

The nouns in Russian are to be classified as animate versus inanimate as the first binary division. Inanimate nouns include not only masculine and feminine divisions but also neuter divisions. The following diagram shows the classification of nouns in Russian.



Natural sex of the object determines the masculine versus feminine division of the animate nouns. In the case of inanimate nouns, the suffix

occuring with the noun stem determines the masculine, feminine and neuter division. Gender clasification under inanimate nouns is grammatical.The suffix of the verb and the suffix of the subject noun must be compatible i.e., the compatibility between the suffixes of the verb and the subject noun forms the basis for the classification of nouns from the level of syntax.

The classification of nouns in Tamil presents a different picture from that of Russian. The major division in Tamil is human Vs non human. Masculine Vs feminine distinction is made under human only and not under non human. The nouns referring to non human animate objects and inanimate objects are grouped under neuter. The table given below will give the differences between Russian and Tamil in the classification of nouns.

NOUNS

RUSSIAN					TAMIL		
ANIMATE		INANIMATE			HUMAN	NON HUMAN	
MAS	FEM	MAS	FEM	NEUT	MAS	FEM	NEUT
MALJ-CHIK	DEVO-CHKA	STOL	KNIGA	OKNO	PAI-YAN	PEN	KAAR

In Russian there are seven cases whereas in Tamil eight cases are found (excluding the traditional vocative in Tamil). A single case is to be equated with two cases in Tamil. In Tamil, genitive and ablative are expressed by separate case forms whereas in Russian one case expresses both genitive and ablative relations. It may be shown as

CASE SYSTEM

RUSSIAN		TAMIL
NOMINATIVE	===========>	NOMINATIVE
ACCUSATIVE	===========>	ACCUSATIVE
INSTRUMENTAL	===========>	INSTRUMENTAL
SOCIATIVE	===========>	SOCIATIVE
DATIVE	===========>	DATIVE
GENITIVE	===========>	ABLATIVE
		GENITIVE
LOCATIVE	===========>	LOCATIVE

Another important difference between Russian and Tamil is the set of forms for expressing each

case relation. In Russian, there are many inflexions for each case: one set is used when the noun is singular and another set when the noun is plural i.e. the number and gender of the nouns determines the case form to be added. The noun determines the case forms to be attached. The case forms are to be classified according to gender and number. For example the plural nouns in Russian take accusative plural form.

In Tamil such a distinction is not found. The noun takes the same case form whether it is singular or plural.

RUSSIAN					TAMIL
ACCUSATIVE					ACCUSATIVE
GENDER			NUMBER		
MAS	FEM	NEUT	SG	PL	

This kind of difference is found in almost all cases. Adjective is found as a grammatical

category in both Russian and Tamil. Adjectives occur in prenominal position in these two languages.

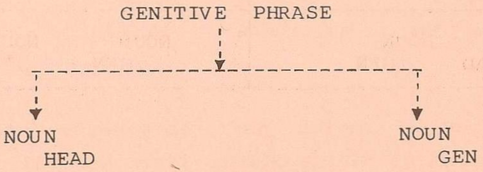
RUSSIAN	TAMIL
ADJ+N	ADJ+N

Though adjectives are found in both the languages and occur in prenominal position, the adjectives in Russian are inflected according to the gender, number and case of the nouns. In other words, the adjective and the headnoun must agree in number, gender and case. But, in Tamil the adjectives are not inflected for number and gender.

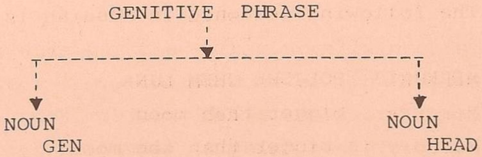
RUSSIAN					TAMIL
ADJECTIVE					CASE ADJECTIVE
GENDER			NUMBER		
MAS	FEM	NEUT	SING	PL	

1.6 WORD ORDER :

We may distinguish levels when we discuss the word order in Russian and Tamil. They are phrase level and sentence level. Let us take the genitive phrase and discuss the differences between Russian and Tamil. In Russian the noun with which the genitive marker is added follows the headnoun. The order may be shown as



In Russian POVERHNOSTJ+MARSG means the surface of the MARS, POVERHNOSTJ is the headnoun and the word with the genitive marker attached follows the headnoun. In Tamil the word with the genitive marker precedes the headnoun. The word order in Tamil is as follows :



The above phrase is translated into Tamil as 'CEVVAAYUTAIYA MEELPARAPPU'.-UTAIYA is the genitive marker in Tamil occurring with the noun 'CEVVAAY'. The headnoun 'MEELPARAPPU' occurs after the word 'CEVVAAYUTAIYA'.

RUSSIAN		TAMIL	
GENITIVE PHRASE		GENITIVE PHRASE	
NOUN HEAD	NOUN GEN	NOUN GEN	NOUN HEAD

Comparative sentences in Tamil and Russian show the differences in word order. CHEM in Russian signals that the sentence is comparative. Similarly 'KAATTILUM' is used in Tamil to signal the comparative sentence. The accusative marker-AI is found occurring with the noun which precedes KAATTILUM .

The following sentence in Russian is comparative.

MERKURIJ BOLJSHE CHEM LUNA

Mercury bigger than moon

'Mercury is bigger than the moon'

Corresponding sentence in Tamil would be

VELLI CANTIRAN-AI KAATTILUM PERITU.
Mercury moon-acc than big
Mercury is bigger than the moon.

By giving the Tamil and Russian sentences one below the other, the difference may be shown as

MERKURIJ	BOLJSHE	CHEM	LUNA
VELLI	PERITU	KAATTILUM	CANTIRAN

The sentence as given in Tamil is ungrammatical. The comparative marker CHEM precedes the noun LUNA in Russian whereas in Tamil the comparative marker 'KAATTILUM' follows the noun 'CANTIRAN' 'moon'. Let us say comparative phrase (CP) consists of a noun and comparative marker.

CP-R =============> com .mar + N

CP-T =============> N + com .mar

In Russian, the verb generally occurs after the subject i.e., the word order in Russian is subject, verb, object (SVO). However we may come across with sentences in which the verb occurs

in other places also. The variation in the occurrence of verb is considered as stylistic. Since the case markers are occurring after the nouns as post positions, the word order may be relatively free. In Tamil, the verb generally occurs in the final position of the sentence. The wordorder in Tamil is not rigid. The wordorder in Russian and Tamil may be shown as

RUSSIAN	TAMIL
SUBJECT+VERB+OBJECT SVO	SUBJECT+OBJECT+VERB SOV

In both the languages the word order is relatively free. The noun phrase which functions as the subject occurs in the initial position of the sentence in both Russian and Tamil. The constituents which occur after V in Russian come before V in Tamil. For example if the structure of a sentence in Russian is S,V,O, LOC, then the sentence structure in Tamil would be S, LOC, O,V. Let us take the following sentence from Russian and Tamil to explain the difference in word order between these two languages.

MALJCHIKI POKUPAJUT1 GAZETYA MAGAZINL
S V O LOC

Boys purchase newspapers shop-in

'The boys purchase newspapers in the shop'

The following is the translation of the above Russian sentence into Tamil.

PAIYANKAL KATAIYIL CEYTITTAALKALAI VAANKUKIRAARKAL
S LOC O V

Boys shop-in newspapers purchase

'The boys purchase newspapers in the shop'

The sentences given below will clarify further the difference between Russian and Tamil in the word order.

R: STUDENTY SLUSHAJUT1 LEKCIJAA PROFESSORG

S V O

Students listen to lecture professor's

'The students listen to the lecture of the professor'

T: MAANAVARKAL PEERACIRIYARUTAIYA VIRIVURAIYAI
S O

KEETKIRAARKAL
V

'The students listen to the lecture of the professor'

R: VRACHI RABOTAJUT1 PROFESSOR I GORODL
S V SOC LOC

Doctors work Professor - with city - in

'The doctors work in the city with the professor'

T: TAAKTARKAL NAKARATTIL PEERACIRIYARUTAN
S LOC SOC

Doctors city in professor with

VEELAICEYK1RAARKAL
V
work

'The doctors work in the city with professor'

1.7 AREA OF PROBLEMS :

When we have one-to-one correspondence in structural comparison, we may not face much problems. One-to-many correspondence poses problems when we are attempting to translate from Russian into Tamil. In order to explain the situation, let us draw a concrete example from Russian and show

how it is handled in our system of translation. In Russian, there is a verb root CHITA *' to study '*

ON CHITAET	he studies
ONA CHITAET	She studies
ONO CHITAET	It studies

The following are the equivalent sentences in Tamil.

IVAN PATIKKIRAAN
he studies

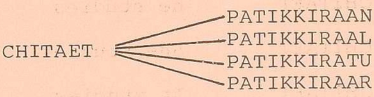
IVAL PATIKKIRAAL
she studies

ITU PATIKKIRATU
it studies

IVAR PATIKKIRAAR
he(Hon) studies

There is only one verb in Russian CHITAET which is to be translated into 'PATIKKIRAAN', PATI-KKIRAAL' PAATIKKIRATU' and 'PATIKKIRAAR' in Tamil. In Tamil we find different markers to denote the person, number and gender of the subject of the

sentence. But in Russian we do not have markers corresponding to the distinctions made in Tamil.



Equivalents were artificially created and the sentences with these distinctions were fed into the Computer. With the verb CHITAET a number was attached to make the distinction. For example CHITAET1 was made as equivalent to 'PATIKKIRAAN' in Tamil. Thus we have the following equivalents.

CHITAET1 ===========> PATIKKIRAAN
CHITAET2 ===========> PATIKKIRAAL
CHITAET3 ===========> PATIKKIRATU
CHITAET4 ===========> PATIKKIRAAR

Many-to-one correspondence between the source language and the target language does not create much problems to the process of translation. In Russian, the adjective is inflected for the gender, number and case of the headnoun. In Tamil the adjective is not inflected for these grammatical aspects. There is no difficulty in

translating the adjective category into Tamil adjective. Translation from Tamil into Russian may pose problems. Similar problems may be faced in the area of semantics. For example if a word in the source language has only one meaning but the corresponding equivalent word in the target language has more than one meaning, then we have to decide the meaning on the bases of both linguistic and non linguistic contexts. Some method has to be devised to establish the equivalents in the meanings of lexical items.

1.8 TRANSFER RULES:

Transfer is one of the components (see the section on components) in the mechanism of translation. It consists of rules which convert the surface structure of Russian into the surface structure of Tamil. In the section that deals with word order, we pointed out the differences between Russian and Tamil in the structure of phrases and sentences.

We observed the noun with the genitive marker occurring after the headnoun in Russian. But in Tamil the noun with genitive markers occurs with the headnoun. By a transfer rule the syntactic structure of Tamil is derived from the

syntactic structure of Russian. We may explain the operation of transfer rule taking an example from Russian and showing how it is translated.

R : PROVERHNOSTJ00 MARS0-G

surface mars'

'the surface of mars'

$\text{Noun}_{\text{HEAD}}$ Noun_{GEN}

The transfer rule takes the $\text{Noun}_{\text{HEAD}} + \text{Noun}_{\text{GEN}}$ as the input and produces the $\text{Noun}_{\text{GEN}} + \text{Noun}_{\text{HEAD}}$ as the output.

R: $\text{Noun}_{\text{HEAD}} + \text{Noun}_{\text{GEN}}$ ---------->

T: $\text{Noun}_{\text{GEN}} + \text{Noun}_{\text{HEAD}}$

It is possible to use the digital number and show the operation of this rule.

$\text{Noun}_{\text{HEAD}} + \text{Noun}_{\text{GEN}}$

1 2 --------->
2 1

As we have noted in the earlier section the comparative sentences also show the difference in these two languages. The comparative marker CHEM in Russian occurs before the noun with which comparison is made. Take the following sentence from Russian.

MERKURIJ	BOLJSHE	CHEM	LUNA
S	ADJ	Com	O
Mercury	big(ger)	than	moon

'Mercury is bigger than the moon'.

VELLI	CANTIRANAI(k)	KAATTILUM	PERITU
S	O	Com	V
Mercury	moon-acc	than	big

'Mercury is bigger than the moon'.

A transfer rule has to move Com from its original place and place it after O.Another rule moves ADJ to the final position. This may be shown as

R:	S	ADJ	Comp	O	
	1	2	3	4	-------->
	1	2	4	3	-------->
T:	1	4	3	2	

Both these rules operate on the surface syntactic structure of the Russian sentence and derive the surface syntactic structure of Tamil sentence.

The verb occurs in the final position of Tamil sentences. We have observed that the verb in Russian follows the subject but precedes the object. A transfer rule is applied to shift the V occurring in the middle position to the final position.

R:	VRACHI	RABOTAJUT-1	PROFESSOR-1	GOROD-L
	S	V	SOC	LOC
	doctors	work	professor-with	city-in

'the doctors work with the professor in the city'.

T: TAAKTARKAL NAKARATTIL PEERACIRIYARUTAN
S LOC SOC
VEELAICEYKIRAARKAL
V
work

'the doctors work with the professor in the city'.

In order to derive the surface syntactic structure of Tamil sentence two transfer rules operate.

R:	S	V	SOC	LOC	
	1	2	3	4	------>
	1	3	4	2	------>
T:	1	4	3	2	

Transfer rules may be considered as a device to derive the Tamil syntactic structure from the syntactic structure of Russian.

The segmentation is here based not on linguistic principles but rather on the basis of the convenience of translation. For example the passive verb in Tamil has been segmented into KAPPATU and -APPATU for the purpose of finding equivalents with the segmentation of Russian passive verb. Perhaps the inclusion of more texts may compel us to revise the program and the method of segmentation.

The various stages of translation from Russian into Tamil may be shown as

R: INPUT — Word order with stylistic variation — Stage: 1

↓

Normal word order — Stage: 2

↓

Tamil word order — Stage: 3

↓

T: OUTPUT — Generation of Tamil sentences — Stage: 4

CHAPTER-2

AN OUTLINE OF RUSSIAN AND TAMIL STRUCTURE DEVISED FOR MACHINE TRANSLATION.

2.1 PROCEDURE :

Any automatic translation system should have the following procedure:

Namely a given text in the source language (say in Russian) has to be translated into the target language (say in Tamil) i.e., the text translated into the target language should contain the same information, that is expressed in the input text. The transformation of the content has to be natural to the maximum extent possible and intelligible as it is expressed in the input text.

A Russian text on astronomy is taken for this experiment. The original Russian text is appended at the end as Appendix-D.

Machine Translation is to be conceived as a practical and operational problem. Therefore, the given text is processed for feeding into the computer.

2.2 FORMATION OF DICTIONARY :

The dictionary is bilingual for the purpose of Machine Translation. Some words in the given text have more than one equivalent in the target language. Since the text chosen is a scientific passage of a particular branch of science, our dictionary shows one-to-one equivalent. This group has not yet worked out any definite procedure to solve polysemy. The problem of polysemy will be taken up at later stage.

It must be noted here that translation of a text means not simply getting equivalents for each word found in the source language. The structure of a given input sentence or phrase has to be analysed, the grammatical categories are to be properly identified and a new sentence or a phrase conveying the sense expressed in it has to be generated in the target language. Therefore, the Machine Translation system should have the entire information about the syntactic structure of sentences to be analysed in source and target language and the relation between them. Besides, the words translated into the target language are to be rearranged properly and a grammatical sentence has to be generated. The algorithm contains the entire set of grammatical rules for the formation of sentences.

It is natural that a researcher as a first step will look into the available bilingual dictionary and grammar of both the source and target languages. In the dictionary for each word there is a grammatical note about the class to which this given word belongs as well as a list of equivalents in the other language. The grammar contains the lists of grammatical categories or class of words and the rules to combine these class of words to form a grammatical sentence. It is assumed that using the grammars and the dictionaries and exactly translating each word, one can write grammatically correct phrases or sentences.

But, unfortunately, none of the dictionaries and grammatical treatises, even the best ones, happens to be adequate, These dictionaries and grammars are written for different purposes. For Machine Translation a different kind of dictionary and grammar is required.

If Machine Translation has to be attempted seriously, a different type of dictionary has to be created. The words have to be classified from different angles, keeping in mind the difficulties that may come across at the time of actual process of Machine Translation. The bilingual grammar comparing the structures of both the languages has to be written. Having worked out all

these details, the process of Machine Translation becomes interesting.

2.3 MORPHOLOGY :

The words found in the text are classified into the following four broad groups: i) Nouns including pronouns and proper names ii) Verbs, iii) Adjectives and iv) Adverbs.

Nouns are entered as they occur in the nominative singular form. Pro-nouns are entered in the nominative as well as in all the case forms.

Verbs are entered in their stem form. The entire dictionary of Russian words in Roman script with Tamil equivalents is given as Appendix-B.

Adjectives are entered in their stem form dropping the gender and number endings.

Adverbs are entered without any change.

2.3.1 NOUNS :

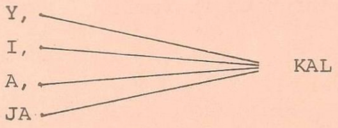
2.3.1.1 SINGULAR VS PLURAL :

Russian singular nouns become plural nouns either by changing the endings (ZAVOD - ZAVODY, DOM - DOMA, GAZETA-GAZETY, KNIGA-KNIGI, PISJMO-PISJMA, POLE-POLJA) or by changing the ending as

well as the stem itself (GRAZHDANIN-GRAZHDANE, REBJONOK-DETI, IMJA-IMENA, SYN-SYNOVJJA).As seen from the above mentioned examples,the plural nouns in Russian have Y, I, A, JA as the most common endings.

In Tamil the suffix 'KAL' is added to make a singular noun into plural irrespective of the gender.

Plural suffix in <u>Russian</u>	Plural suffix in <u>Tamil</u>



Keeping in mind the limited memory of the Computer instead of providing exhaustive rules to convert singular into plural, the following simple technique is adopted for this purpose:

A marker 'X' is added to the Russian singular nouns to make them plural, where 'X' is a plural marker.

Examples :

Singular		Plural	
Russian	Tamil	Russian	Tamil
PLANETA	KOOL	PLANETAX	KOOLKAL
GAZ	VAAYU	GAZX	VAAYUKKAL
ATMOSFERA	KAARRUMANTALAM	ATMOSFERAX	KAARRUMANTAL-ANKAL

In the actual process of translation the marker 'X' used in the Russian word is identified as plural marker, and the suffix 'KAL' is added to the singular noun in Tamil.

2.3.1.2 CASES :

Russian nouns, pronouns, adjectives and numerals can be declined; that is, they change according to cases and numbers. For example: the noun 'STOL' has this form in the sentence STOL ZDESJ, but another form (STOLE) in the sentence 'KNIGA NA STOLE'. The form 'STOL' is the nominative case of the singular and 'STOLE' the prepositional of the same number. In plural it will be 'STOLY' and 'STOLAH' in nominative and prepositional cases respectively.

The Russian system of declension is constituted by the following cases:

1) Nominative
2) Genitive
3) Dative
4) Accusative
5) Instrumental
6) Prepositional

and two numbers :

1) Singular and
2) Plural

Let us take accusative case of nouns in Russian and compare them with the accusative case in Tamil.

The endings of the accusative case of nouns are -U and -JU used in the singular of those nouns which have -'A' or -'JA' in the Nominative singular, the -U corresponding to the -A ,and the -JU to the -JA. Otherwise the accusative is like either the Nominative or the Genitive of the same number.

(a) Accusative singular = Nominative singular in:

(1) Masculine nouns denoting inanimate objects

also with collective nouns, such as Narod, Klass thus:

STOL - STOL
TRAMVAJ - TRAMVAJ
PORTFELJ - PORTFELJ

(2) Feminine nouns ending in -J in the Nominative singular as DVERJ-DVERJ,TETRADJ-TETRADJ, LOSHADJ-LOSHADJ, DOCHJ-DOCHJ, etc.

(3) All neuter nouns:OKNO-OKNO,MESTO-MESTO,ZDANIE-ZDANIE, etc.

(b) Accusative plural = Nominative plural in all types of nouns (masc., fem; neut) denoting inanimate objects; thus: (masc) STOL-STOLY, JASHCHIK-JASHCHIKI, TRAMVRAJ-TRAMVAJ, PORTFELLJ - PORTFELJ etc., (fem) STENA-STENY, KNIGA-KNIGI, CHITALJNJA-CHITALJNI,AUDITORIJA-AUDITORIJ,TETRADJ-TETRADI, DVERJ-DVERI, etc. (neut)OKNO-OKNA, MESTO-MESTA, ZDANIE-ZDAMIJA, so also in collective nouns as (fem) SEMJJA-SEMJI.

(c) Accusative singular = Genitive singular in masculine nouns denoting animate beings, that is, persons and animals (beasts, birds, etc.) if

the Nominative singular has no case-ending, as in STUDENT-STUDENTA, INZHENER-INZHENERA, BRAT-BRATA, etc, (but ILJJA, masculine proper name ILYJJU, as the Nominative singular ends in -JA).

(d) Accusative plural = Genitive plural in both masculine and feminine nouns denoting animate beings: (masc)

STUDENT-STUDENTOV, INZHENER-INZHENEROV, TOVARISHCH-TOVARISHCHEJ, BRAT-BRATJEV, ANGLICHANIN-ANGLICHAN, etc.,

(fem) STUDENTKA-STUDENTOK,
AMERIKANKA-AMERIKANOK,
ZHENA-ZHJON,
LOSHADJ-LOSHADEJ,
KOROVA-KOROV.

This is the complex nature of case-system in Russian. The complexity of case-system in Russian is due to two factors: (i) Firstly there are several endings for each case. (ii) Secondly, in some words, the endings of a particular case is the same as the endings of other cases.

Therefore, an exhaustive analysis of case-systems in Russian and Tamil has to be done for the purpose of Machine Translation. In Russian, unlike in Tamil, the case endings even for a particular case may be many. Depending on the type of noun, the case endings differ. There are three types of nouns namely masculine, feminine and neuter.

Besides, the case endings will differ depending on whether the noun is animate or inanimate. These three types of classifications with their Tamil equivalent are given in Appendix-C.

In the absence of a detailed analysis of case-systems in Russian and Tamil and considering the fact that the computer has limited memory, the research group was forced to find a way out to tackle this complex problem.

It was decided to replace the various case inflexions by case markers. The following case markers are suffixed to denote the case endings :

Marker	The Russian case for which the marker represents	Tamil case endings
O	Nominative	No change
A	accusative	AI
G	Genitive	UTAIYA
R	Ablative	ILIRUNTU
I	Instrumental	AAL
D	Dative	UKKU
L	Locative	IL
S	Associative	UTAN
M	Locative (Human)	ITAM

Let us explain here briefly by a few examples how the case-endings are replaced by case-markers, and how the case markers are identified and the translation is effected.

Example 1

Original: V SOLNECHNOJ SISTEME DEVJATJ PLANET:

The underlined phrase is in the locative case. In the input text this phrase in the locative case is replaced by case-marker L. The input text is pre-edited as shown below.

Input: SOLN*SISTEMA0L DEVJATJ00 PLANETAX0.

The adjective 'SOLNECHNOJ' is reduced to the stem SOLN and joined to the noun SISTEMA by a link-marker *. The noun 'SISTEME' is in the locative case singular. Therefore, it is replaced by SISTEMA0L. Here 0 denotes singular and L locative case. The numerals 'DEVJATJ' is in singular nominative. Therefore it is fed as 'DEVJATJ00'.

In Russian after the cardinal numerals 1 to 4 the noun is put in the genitive singular form and after 5 to 20 the noun is put in the genitive plural. Since such a system is not prevalent in Tamil, <u>after the cardinal numbers, nominative case marker 0 is added instead of the genitive case in Russian</u>.

The computer is programmed to identify the case-markers and generate the Tamil sentences. As an output text we get the following.

<u>Output</u> : CUURIYA MANTALATTIL ONPATU KOOLKAL.

For the Russian word 'SISTEMA' the dictionary gives the meaning 'MANTALAM'. SISTEMA0L → 'MANTALAMIL'.

How 'MANTALAMIL' is edited to 'MANTALATTIL'? See programming techniques.

<u>Example 2.</u>

Original --> RAZMERY, MASSA I PLOTNOSTJ VENERY
BLIZKI K ZEMNLY.

<u>Input --></u> <u>RAZMERAX0<MASSA00<PLOTNOSTJ00<VENERA0G</u>
@BLIZKJUT2 ZEMN*TIP00 PLANETAXA.

The underlined genitive phrase is taken as one unit. The last word in this phrase VENERA is in genitive case. While translating into Tamil it has to be transferred to the beginning of the phrase. Therefore, the marker '<' is attached so that it can trace its place in the target language. The computer is programmed to pre-edit the input text automatically and rearrange the phrase in the following way.

<u>Pre-edited:</u> VENERA0G RAZMERAX0 MASSA00 PLOTNOSTJ00

<u>Output</u> : VELLIUTAIYA ATARVU PARIMAANAM ETAI

Example 3.

Original : GLUBINA GAZOVOGO SLOJA NA PLANETAH-
GIGANTAH OCHENJ VELIKA

Input : GLUBINA00<GAZOV*SLOJOG
GIGANTY*PLANETAXL OCHENJ @VELIKET3.

The underlined genitive phrase is pre-edited by the computer automatically and rearranged in the following way.

Pre-edited : GAZOV*SLOJOG GLUBINA00
GIGANTY*PLANETAXL OCHENJ @VELIIKET3.

Output : VAAYU ATUKKUTAIYA AALAM PERIYA
KOOLKALIL MIKA ATIKAMAAYIRUKKIRATU.

Example 4.

Original : OCHENJ NIZKAJA TEMPERATURA
ATMOSFER PLANET-GIGANTOV.

Input : OCHENJ00<NIZK*TEMPERATURA00<ATMOSFERAXG
GIGANTY*PLANETAXL.

Pre-edited : ATMOSFERAXG OCHENJ00
NIZK*TEMPERATURA0V
GIGANTY*PLANETAXL.

MIKA PERIYA KOOLKALIL
KAARRUMANTALANKALUTAIYA KURAINTA
VEPPANILAI

Example 5.

Original: VNUTRENNIE SLOI ATMOSFERY NA JUPITERE
ISPYTYVAJUT OGROMNOE DAVLENIE VERHNIX
SLOJOV GAZOV.

Input : VNUTREN*SLOJX0<ATMOSFERAXG<JUPITER0L
@ISPYTYVAJUT2
OGROMN*DAVLENIE00<VERHN*SLOJXR<GAZXN.

Pre-edited: JUPITOR0L ATMOSFERAXG VNUTREN*SLOJXO
@ISPYTYVAJUT2 GAZXN VERXN*SLOJXR
OGROMN*DAVLENIE00.

Output : VIYAALANIL KARRUMANTALATTUTAIYA UL
ATUKKUKAL VAAYUKALIN VELI
ATUKKUKALILIRUNTU ATIKA ALUTTAM
PERUKINRANA.

Example 6.

Original. CHETYRE GLAVNYH SPUTNIKA JUPITERA

Input : CHETYRE00<GLAVN*SPUTNIKX0<JUPITER0G

The word JUPITER in the genitive case has to be brought before 'CHETYRE'. Therfore, the three components in the genitive phrase are linked together by the marker '<'. The computer identifies this marker and rearranges the phrase for automatic translation. The rearranged pre-edited phrase will take the following shape.

JUPITEROG CHETYREOO GLAVN*SPUTNIKXO

As it was explained in example 1, after the cardinal number 'CHETYRE' nominative case marker O is added instead of the genitive case in Russian.

Output: VIYAALANUTAIYA MUKKIYA TUNAIKKOOLKAL NAANKU

How these case-markers are identified and replaced by Tamil case endings is explained in the chapter on programming techniques.

2.3.2 ADJECTIVES :

In Russian, the adjective changes according to gender, number and case of the noun. The adjective qualifies its head-noun and agrees with it, i.e., it agrees with the gender, number and case of the head-noun. Since, the adjectives in Tamil

do not change for gender, number and case of the head-noun, the adjectives are entered in the dictionary in the stem form.

In both the languages the adjective qualifies the noun and they occur together, the adjective is linked to the noun by a link-marker '*'. During the translation process the adjective and noun combinations are taken as a single unit and the case-marker is attached only to the head-noun. The following examples illustrate the process of translation.

SOLN*SISTEMA0L--> CUURIYA MANTALATTIL
DRUG*PLANETAXA -> MARRA KOOLKALAI
BEL*SHAPKAX0 --> VELLAI MUKALKAL
NIZ*TEMPERATURA00 -> KURAINTA VEPPANILAI

2.3.3 VERBS :

The Russian verb has the following forms: the infinitive (CHITATJ), the indicative mood (the present tense CHITAJU), the past tense (-CHITAL), the future tense (-Budu CHITATJ), the imperative (CHITAJ) and the conditional mood (CHITAL BY).

The Russian infinitive names an action; CHITATJ-to read, NESTI-to carry BERECHJ-to take of, and it does not denote the person or tense.

The infinitive of a verb ends in the suffix-TJ (CHITATJ),TI (NESTI) or CHJ (BERECHJ).

To obtain the stem of the infinitive, the infinitive suffix must be dropped, for example CHITA-(TJ), STROI(TJ), SMOTRE(TJ), TJANU(TJ), NES-(TI), VEZ(TI), RAS(TI).

The stem of the infinitive is used to form the past tense of the verb (CHITA-L,NJOS,VJOZ,ROS), the past participle (CHITA-VSHIJ, NJOS-SHIJ,VJOZ-SHIJ, ROS-SHIJ) and the adverbs of perfective verbs (PROCHITA-V,PRINJOS-SHI or PRINESJA,PRIVJOZ-SHI).

Verbs whose infinitive ends in -CHJ namely BERECHJ, STERECHJ, MOCHJ, PECHJ etc. developed historically from the forms BEREGTI, STEREGTI , MOGTI, PEKTI and all their forms are obtained from a stem ending in G or K, i.e., from the stems BEREG-,STEREG-,MOG-,PEK.

As stated earlier, the verbs are stored in the stem form for the above mentioned reasons.All other forms are added to the stem to make it present, past or future tense.

2.3.3.1 PRESENT TENSE :

In Russian the verbs are classified into several groups. But for the purpose of Machine Translation all the Russian verbs are classified into two major groups. According to their personal endings, verbs fall into two groups. Verbs of the first conjugation and verbs of the second conjugation.

Verbs of the first conjugation are marked as @ group. They are conjugated for the present tense as shown below.

Russian	Tamil
CHITA	PATI
JA CHITAJU	NAAN PATIKKIREEN
TY CHITAESHJ	NII PATIKKIRAAI
ON1 CHITAET1	IVAN PATIKKIRAAN
ONA1 CHITAET2	IVAL PATIKKIRAAL
ONO CHITAET3	ITU PATIKKIRATU
ON4 CHITAET4	IVAR PATIKKIRAAR
MY CHITAEM	NAANKAL PATIKKIROOM
BY CHITAETE	NIINKAL PATIKKIRIIRKAL
ONI1 CHITAJUT1	IVARKAL PATIKKIRAARKAL
ONI2 CHITAJUT2	IVAI PATIKKINRANA

Verbs of the second conjugation are marked as 0 group. This type of verbs are conjugated for the present tense as shown below:

Russian	Tamil
OSTROI	KATTU
JA STROJU	NAAN KATTUKIREEN
TY STROISHJ	NII KATTUKIRAAI
ON1 STROIT1	IVAN KATTUKIRAAN
ONA1 STROIT2	IVAL KATTUKIRAAL
ONO STROIT3	ITU KATTUKIRATU
ON4 STROIT4	IVAR KATTUKIRAAR
MY STROIM1	NAANKAL KATTUKIROOM
VY STROITE	NIINKAL KATTUKIRIIRKAL
ONI1 STROJAT5	IVARKAL KATTUKIRAARKAL
ONI2 STROJAT6	IVAI KATTUKINRANA

There are many other verbs whose conjugations do not fall within these two types of verbs. For the time being the Russian verbs are to be identified with one of these two broad groups to avoid programming complications.

2.3.3.2 PAST TENSE :

The past tense for the two groups are formed by adding past tense endings L, LA, LO, LI to the stem of the verbs. To distinguish the different endings in Tamil, the following pattern is adopted.

JA CHITAL3	NAAN PATITTEEN
TY CHITAL4	NII PATITTAAY
ON1 CHITAL1	IVAN PATITTAAN
ON4 CHITAL2	IVAR PATITTAAR
ONO CHITALO	ITU PATITTATU
ONA1 CHITALA	IVAL PATITTAAL
MY CHITALI2	NAANKAL PATITTOOM
VY CHITALI3	NIINKAL PATITTIIRKAL
ONI1 CHITALI1	IVARKAL PATITTAARKAL
ONI2 CHITALI4	IVAI PATITTANA

The same pattern is adopted for the other groups of verbs also.

How these present tense and past tense markers are identified and the synthesis of conjugation in Tamil verb stems takes place are explained in the programming techniques.

2.3.3.3 FUTURE TENSE :

In Russian there are two types of future tenses (1) perfect future and (2) imperfect future. The perfect future is known as simple future. Its endings resemble those of the present tense form of the imperfective verbs and the imperfect future is known as compound future. It is formed by means of the future tense of the auxiliary verb BYTJ and the infinitive of the imperfective verb.

The group has not yet attempted to find a method to program this aspect. This will be attempted in phase II.

In Russian, the verb can be placed in any position whereas in Tamil the verb is generally placed at the end of the sentence. First of all the verb in the Russian sentence has to be identified and shifted to the final position automatically for Machine Translation.

2.3.4 ADVERB :

The adverb is an invariable part of speech. In a sentence, adverb generally modifies a verb: RABOTATJ VMESTE. An adverb may also modify (a) an adjective or (b) an adverb or (c) a noun.

According to their meaning, adverbs fall into the following main groups. (1) adverbs of manner (2) adverbs of time (3) adverbs of place (4) adverbs of measure and degree (5) adverbs of purpose (6) adverbs of cause.

Out of these groups, the adverbs of manner alone is treated here as adverb and others are treated as nouns. As the adverb of manner modifies the verb, the adverb and verb are linked by a marker '*' This word combination is treated as a single word for the purpose of Machine Translation.

2.4 SYNTAX :

2.4.1 SENTENCE PATTERN :

The sentences are classified into two broad patterns: (1) verbal and (2) non-verbal.The non-verbal sentences are identified and seperated from verbal ones by providing a marker # at the beginning of the sentence.

In the verbal sentences the verb is identified by the verb classification markers @ and 0.

These markers serve two purposes. First it serves to identify the verb and to shift the verb to the final position. Secondly they provide the possibility to choose one of the two types of conjugation included in the programme.

2.4.1.1 ACTIVE VS PASSIVE :

Let us compare active and passive constructions in both the languages.

Russian	Tamil
Active.	
RABOCHIE STROJAT DOM:	TOLILAALIKAL VIITU KATTUKIRAARKAL
Passive.	
DOM STROITSJA RABOCHIMI:	VIITU TOLILAALIKALAAL KATTAPPATUKIRATU.

In Russian the reflexive participle -SJA denotes the passive character of the verb.

The computer can be programmed to identify the reflexive participle. For simplicity the passive construction is marked by the symbol'*'. Therefore, the sentence DOM STROITSJA RABOCHIMI is fed to the computer as DOM STROIT* RABOCHIJXI.

Example 1.

Original. ATMOSFERA DELITSJA OBLAKAMI NA DVA SLOJA.

Input. ATMOSFERA00 0DELIT3* OBLAKOXI DVA00 SLOJXK.

Output. KAARRUMANTALAM MEEKAMKALAAL,IRANTU ATUKKUKALAAKA PIRIKKAPPATUKIRATU.

Example 2.

Original. POVERHNOSTJ MARSA NABLJUDAETSJA UCHJONYMI.

Input. POVERHNOSTJ00<MARSOG @NABLJUDAET*+ UCHJONIJXI.

Output. CEVVAAYUTAIYA MEELPARAPPU VINGAANIKALAAL URRUNOOKKAPPATUKIRATU.

While translating passive constructions into Tamil, the classification of Tamil verbs into weak and strong groups has to be kept in mind. The weak verbs are marked as '+' and the strong verbs are marked as '-'. This marking can be done in the dictionary of the target language. But at present the marking is done in the pre-editing stage.

2.4.1.2 PASSIVE SENTENCES WITH SHORT PARTICIPLES AND ADJECTIVES:

Just like reflexive verbs short form of participle is used to express passive meaning. Present and past participles are formed from the stem of the infinitive or past tense by means of the suffixes. The participles can be identified by their suffixes.

In Russian science texts, one often comes across, short form of participles and adjectives used as predicate. In such cases the participles and adjectives losing their general function of qualifying the noun play the role of a predicate. Since such usage is not prevalent in modern Tamil, the short form of Russian participles and adjectives are treated as normal verbs. Verbal suffixes are added to these participles and adjectives.

Therefore, the input text is converted into Meta-language-1. This Meta-language-1 is the coded language for the easy recognition of the grammatical categories. In a main frame Computer by providing exhaustive table and grammatical rules the difference between the natural language and Meta-language can be reduced to a minimum. Till then the input text shall not be in the natural language.

Let us take a few examples and show how the grammatical categories of the natural language are converted to Meta-language and the Machine translation is effected.

Original: PLUTON BYL OTKRYT TOLJKO V 1930 GODU

'OTKRYT'- is the short form of passive participle. 'BYL' denotes that the passive participle is in the past tense.

In the bilingual dictionary the Tamil equivalent for the stem of the verbs 'OTKRY' is given as 'KANTUPITI'. The past tense of the auxiliary verb and the passive participle ending 'T' have to be recognised by the Computer. In the present system of translation the passive past participle is provided with a marker '&'. The Tamil equivalent can be a weak or a strong verb. As in the case of active and passive construction '+' is added for weak verb and '-' is added for strong verb.

Therefore, the passive participle construction mentioned above shall be converted to a Meta language-1 as shown below.

Input: PLUTON OTKRYS&+ TOLJKO<1930XL.

Hence the symbol 'S' stands for singular.

Output: PULUTOON 1930 KALIL TAAN
KANTUPITIKKAPPATTULLATU.

How the markers are identified and the meaningful sentence is generated in Tamil is explained in the programming techniques.

2.4.1.3 A RUSSIAN PASSIVE SENTENCE WITH SHORT FORM OF ADJECTIVE:

Let us take an example for a passive construction with short form of adjective:

RAZMERA, MASSA, PLOTNOSTJ VENERY BLIZKI K ZEMNLI

'BLIZKI' is the short form of the adjective BLIZKIJ. Since the subject is in plural I is added to the short adjective BLIZK. Now in our bilingual dictionary 'BLIZK' is entered in the adjective class of words. But here 'BLIZK' is used as a predicate. Hence adjectives used as predicate have also to be listed in the verb class of words. Of course, this is a duplication of words in the dictionary. But at the present stage this is unavoidable. A solution has to be found at a later stage. It is a problem for future research.

Once the adjective finds a place in the dictionary with verb group, then personal endings can be attached to it and treated on par with verbs.

Therefore, the example mentioned above is converted to Meta-language-1.

Input: RAZMERA MASSA<PLOTNOSTJ<VENERA0G
BLIZKJUT2 ZEMN*TIP PLANETA0A

The short form of adjective is converted to a verb stem and the verbal suffix for the generation of present tense is attached to it. The computer analyses this as a verbal structure.

A few more examples are given to explain this process of conversion and identification for Machine Translation.

Example-1.

Original : VENERA OKRUZHENA ATMOSFEROJ

Input : VENERA OKRUZHIS&+ ATMOSFERA0I.

Output : VELLI KAARRUMANTALATTAAL
CUULAPPATTULLATU.

Example-2.

Original : S ZEMLI HOROSHO VIDNY BELYE SHAPKI POLJUSOV.

Input : ZEMLJA0R ØHOROSHO*VIDIAT6 BEL*SHAPKAX0<POLJUSXG

Output : PUUMIYILIRUNTU TURUVANKALUTAIYA VELLAI MUKALKAL NANRAAKA PAARKKAPPATUKINRANA.

Original : ATMOSFERY OBNARUZHENY NA PLANETAH GIGANTAH.

Input : ATMOSFERAX0 OBNARUZHIP& GIGANIY*PLANETAXL.

Output : KAARRUMANTALANKAL MIKAPPERIYA KOOLKALIL KANTUPITIKKAPPATTULLANA.

Example-3.

Original : GLUBINA GAZOVOGA SLOJA NA PLANETAH GIGANTAH OCHENJ VELIKA

Input : GLUBINA00<GAZOV*SLOJ0G GIGANTY*PLANETAXL OCHENJ00 @VELIKET3

Output : VAAYU ATUKKUTAIYA AALAM PERIYA KOOLKALIL MIKA ATIKAMAAYIRUKKIRATU.

Example-4.

Original : POVERHNOSTJ ETIH PLANET ZAKRYTA OBLAKAMI

Input : POVERHNOSTJ00<PLANETAXG ZAKRYS&+ OBLAKOXI

Output : KOOLKALUTAIYA MEELPARAPPU MEEKANKALAAL MUUTAPPATTULLATU.

Example-5.

Original : U JUPITERA IZVESTNO 12 SPUTNIKOV.

Input : JUPITEROL ESAT6 12 SPUTNIKX0.

Output : VIYAALANIL 12 TUNAIKKOOLKAL IRUKKINRANA.

2.4.2 DEGREES OF COMPARISON :

Adverbs formed from qualitative adjectives have degrees of comparison. The comparative degree of adverbs has the same form as that of adjectives:

Let us take the following Russian sentence with comparative degree.

MERKURIJ	NEMNOGO	BOLJSHE	CHEM	LUNA
1	2	3	4	5

The same sentence can also be written as

MERKURIJ	NEMNOGO	BOLJSHE	LUNY
1	2	3	5

But, in this phase a sentence with the word 'CHEM' alone is attempted for translation. The presence of the word 'CHEM' serves as an indicator for the Computer to identify this type of construction.

Though it was mentioned earlier that the word-order in Russian and Tamil is not so rigid, certain phrases in the sentence have fixed word order. If we change the word order, the sentence seems to be unnatural. Hence we have to find out a generally acceptable form which may be the style for the purpose of Machine Translation of scientific and technical literature.

The literal word-to-word translation of the above mentioned sentence will be:

PUTAN	KONCAM	PERITU	KAATTILUM	CANTIRANAI
1	2	3	4	5

But in Tamil this sentence is ungrammatical and unacceptable. The generally acceptable from will be

PUTAN CANTIRANAI KAATTILUM KONCAM PERITU
1 5 4 2 3

In order to rearrange the structure to a generally acceptableform, a definite rule has to be provided in the algorithm.

The computer recognises the participle "CHEM". The presence of the participle "CHEM" is the indicator for a comparative sentence.Whatever that precedes the comparative participle 'CHEM' is taken as the comparative phrase. This phrase is automatically placed by the computer at a position immediately after the subject to which the former is compared. The following example will illustrate the restructuring and the translation process involved.

Source Language: 1

1	2	3	4	5
KLIMAT	NA MARSE	HOLODNEE	CHEM	NA ZEMLE

Input: Meta Language-1:

1	2	3	4	5
KLIMAT00	MARSOL	HOLODNEE00	CHEM00	ZEMLJA0A

Meta Language-2 :

1 5 4 2 3
KLIMATOO ZEMLJAOA CHEMOO MARSOL HOLODNEEOO

Target Language :

1 5 4 2,
TATPAVEPPAM PUUMIYAI KAATTILUM CEVVAAYIL

3
ATIKAKULIR

The meaning of the phrase to be translated is fixed with the help of a special semantic language known as Meta-language-2. Its structure serves as a convenient tool to transfer the meaning while translating from Russian to Tamil.

Meta language is not an universal intermediate language structured on logic or logical semantics. It is based on a principle of maximum similarity to the natural language. The input into the Meta language-2 should have the same form as the content of the phrase occuring in the natural language. The syntactic means of all the Meta languages are similar and coincide with the features of deep structure of natural languages.

But on the other hand the lexical feature of the Meta languages carry a native character. Therefore, Target language (Tamil) is like a preparatory stage of natural Tamil. The elements segmented in it preserve their own Tamil structure. While designing a basic language certain identification for the rearrangement of the word can be given.

2.5 AUTOMATIC SYNTHESIS OF TEXTS IN A NATURAL LANGUAGE:

Let us take a Russian sentence with a genitive phrase.

RAZMERY, MASSA, PLOTNOSTJ VENERY BLIZKI K ZEMNLI In the above sentence 'VENERA' is used in the genitive case in the phrase.

RAZMERY,	MASSA,	PLOTNOSTJ,	VENERY
1	2	3	4

While translating this phrase into Tamil the arrangement of the words has to be altered. The acceptable order in Tamil shall be 4,1,2,3,. Therefore, the computer has to identify the genitive marker in the phrase and the word with the genitive marker has to be brought to the beginning of the genitive phrase. Some procedure has to be adopted so that the genitive phrase has to be

demarked in the sentence and rearrangement of the genitive phrase is effected automatically. Several alternatives were discussed and tried. Finally, an arrow is used to identify the genitive phrase. For Example,

RAZMERY<MASSA<PLOTNOSTJ<VENERY.

This input phrase will be converted to Meta language-1 by the Computer automatically.

VENERA0G RAZMERAX0, MASSA00 PLOTNOSTJ00.

This will be translated into Tamil as

VELLIUTAIYA PARIMAANAM ETAI ATARVU.

The following examples shall demonstrate the process of translation of a Russian genitive phrase:

POVERHNOSTJ MARSA.

Input: POVERHNOSTJ00<MARS0G

Meta Language 1: MARS0G POVERHNOSTJ00.

Meta Language 2: CEVVAAYUTAIYA MEELPARAPPU.

In the following Russian sentence there are three genitive phrase. Let us see how the sentence can be fed into the Computer and how the translation is effected.

Original : S ZEMLI HOROSHO VIDNY BELYE SHAPKI
POLJUSOV TJOMNYE OBLASTI MOREJ,
LINII KANALOV.

Input : ZEMLJA0R HOROSHO*VIDJAT6+ BEL*SHAPKAX0
<POLJUSXG TJOMN*OBLASTX0
MOREXG<LINIJAX0<KANALXG

Meta Language 1 : ZEMLJA0R HOROSHO*VIDJAT6+
POLJUSXG BEL*SHAPKAX0 MOREXG
TJOMN*OBLASTX0<KANALXG LINIJAX0·

Meta Language 2 : PUUMIYILIRUNTU NANRAAKA
PAARKKAPPATUKINRANA TURUVAMKALUTAIYA
VELLAI MUKALKAL SAMAVELIUTAIYA IRUNTA
PIRATEECAMKAL, KAALVAAYKALUTAIYA
VARIKAL.

Output : PUUMIYILIRUNTU, TURUVAMKALUTAIYA
VELLAI MUKALKAL, SAMAVELIUTAIYA
IRUNTA PIRATEECAMKAL, KAALVAAYAKALTA-
IYA VARIKAL NANRAAKA PAARKKAPPATUKIN-
RANA.

Here again certain amount of refinement is still required to make it more idiomatic. In idiomatic Tamil it will read like,

PUUMIYILIRUNTU, TURUVAMKALUTAIYA VELLAI MUKALKALUM, SAMAVELIUTAIYA IRUNTA PIRATEECAMKALUM KAALVAAYKALUTAIYA VARIKALUM NANRAAKA PAARKKAPPATUKINRANA.

When there is a conjunction between two sentences or between two phrases, it is observed that different means are used in Tamil to join them. A module has yet to be worked out for this problem. It shall be dealt in phase II.

2.6 COMPLEX SENTENCES :

So far we were discussing about simple sentences. Of course we have not yet analysed all the features of simple sentences. As it was mentioned earlier that future tense, infinitive constructions are not yet analysed in this phase. Similarly certain other features also remain beyond the scope of this work.

However, there are two complex sentences in this text. Any compound or complex sentence can be converted into a simple structure. A complex

sentence is taken for analysis and its results are given below:

Let us, for example, take the following Russian sentence and analyse:

OCHENJ NIZKAJA TEMPERATURA ATMOSFERY NA GIGANTAH PLANETAH POKAZYVAET, CHTO ONI POLUCHAJUT TEPLO OT SOLNCA,

It is easy to see that the direct translation i.e., word by word literal translation of the given syntactic structure is meaningless:

* MIKA KURAINTA VEPPANILAI KAARRUMANTALATIN-UTAIYA PERIYA KOOLKALIL TERIVIPPATU,ENRU IVAI PERUKINRANA VEPPAM CUURIYANILIRUNTU.

Though many readers may correctly guess the meaning from the Tamil translation it cannot any case be accepted. Let us rearrange Tamil translation by splitting the complex sentence into two simple sentences.

GIGANTY-PLANETY POLUCHAJUT2 TEPLO IZ SOLNCA. POETOMY TEMPERATURA ATMOSFERY V GIGANTAHPLANETAH OCHENJ NIZKAJA.

The Computer can analyse these two simple structures and give us the following translation.

CUURIYANITAMIRUNTU PERIYA KOOLKAL VEPPAM PERUKINRANA. ENAVEE KAARRUMANTALATINUTAIYA VEPPAM PERIYA KOOLKALIL MIKA KURAIVAAKA IRUKKIRATU.

This way of splitting the compound and complex sentences into simple sentences can be done, till a module for the former is worked out. However, an attempt was made to analyse this complex sentence as detailed below:

OCHENJ<NIZK*TEMPERATURA00<PLANETAXG GIGANTY*PLANETAXL POKAZYVAET3 CHTO100 ONI2 POLUCHAJUT2 TEPL00<SOLNCE0R.

This type of complex structure is identified by the conjunction CHT01. While translating into Tamil, the subordinate clause is first translated. Then the conjunction is placed. Finally the main clause is attempted.

An algorithm is provided to this effect and the text is pre-edited automatically and the translation is effected.

The following output was obtained:

PERIYA KOOLKAL CUURIYANILIRUNTU VEPPAM PERUKINRANA ENRU KAARRUMANTALAMKALUTAIYA KURAINTA VEPPANILAI MIKA PERIYA KOOLKALIL TERIVIKKIRATU.

We have not yet analysed all the possible types of complex structures. Therefore, it is too early to make a final statement that all the complex sentences can be tackled in this module. Exhaustive analysis of various complex structures shall be attempted in phase-II.

The input text is appended as Appendix-D. The pre-edited text is given as Appendix-F. The translation in Roman alphabets is given as Appendix-G. The Tamil version is given as Appendix-H.

In conclusion it can be said that the Machine Translation is attempted here as a practical problem. The natural chracteristics of both source and target languages are maintained to a maximum possible extent. In a main frame Computer, by providing case inflexion table, the case markers can be removed and the difference between the original and input text can be further minimised.

CHAPTER- 3

PROGRAMMING ASPECTS OF MACHINE TRANSLATION SYSTEM

INTRODUCTION

There are various techniques by which a Computer can be of help for analysis and study of the syntax of a language and Translation. Generally speaking translation is an art of handling and manipulating the lexical power of linguistic elements of a language. Hence we can call the translation procedure as an intellectual exercise involving the expertise of a language. and its grammar.

A translator should combine sensitivity with intelligence, creativity with good organization and inventiveness with discipline. Hence a human translator must not only be a good scholar in both source and target languages but also he should have the ability and skill to deal with the most difficult situation during the process of translation. So, for a machine, to carry out the translation it will have to be taught source and Target languages and also the set of translation rules.

The Translation process can be classified as follows:

a) Human Translation -- HT
b) Machine aided human Translation -- MAHT
c) Human aided machine Translation -- HAMT
   (EX: Systems like LOGOS, TITUS)
d) Pure Machine Translation -- MT

This paper discusses the functional organization of the Human aided Machine Translation System (HAMT). Human aided Machine Translation System can be adopted to translate Technical papers, equipment manuals, laboratory procedures, technical and Scientific articles, military records and weather reports etc.

With the recognition that Computers offer a great potential for assistance in all areas of human endeavour, it is also possible to assist in the task of natural language processing and automatic translation.

The earlier chapters were devoted for discussing the way the project was initiated and the history of Machine Translation and the general grammatical pattern, the source and target

language structure and different methods devised for effecting the machine translation system etc.

This chapter is devoted to explain the technical and programming intricacies as how the actual system is programmed and implemented on the Computer.

Before proceeding to explain the technical aspect of the Russian to Tamil Machine Translation system called as TUMTS (Tamil University Machine Translation System) it might be useful to understand certain concepts.

* Machine Translation is considered primarily an operational rather than a theoritical system.
* Russian is considered to be a *"highly inflectional language"*.
* A mathematical representation of a language and its syntax is very much essential for an effective and efficient processing utilising a Computer.

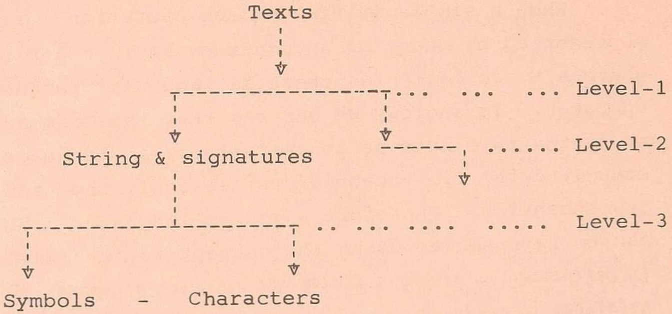
* Certainly humans do not find all the syntactic interpretations of a natural language sentences, before they consider the meaning of the sentence, whereas the exhaustive syntactical interpretation is possible with the help of a Computer.

* When we say that the translation process is nothing but essentially creating a correspondence between a source and target language, it is felt necessary to define source and target languages with the help of a simple mathematical forms.

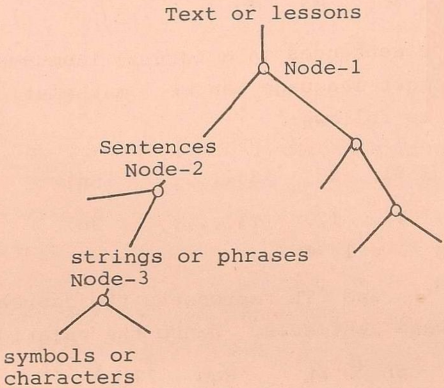
* *"It is assumed that a given number of lessons or texts is supposed to define a particular degree of language competance, the acquisition of which is analysed into a number of steps or learning process".*

Therefore in a language the constituents without actually violating the syntax, can be represented as follows,

Consider a natural language which is hierarchically oriented with a 'TOP DOWN' structure representation.



By Binary tree notation.



When a simple multiplication operation is represented by means of an equation say $M = A \times B$, where A, B are called as operands and 'X' is the operator. In anology we can say that the Grammar or syntax rule acts as an operator on a language constituents i.e vocabulary or lexical item and its semantics. Therefore a source language SL can be represented using the concept as "*a text by definition is simply a finite set in which a number of relations is defined*"

$$SL = (\alpha_1, \alpha_2, \alpha_3, \ldots\ldots\ldots \alpha_n)$$

Where $\alpha_1, \alpha_2, \ldots\ldots \alpha_n$ are a set of texts of the source language conforming to the syntax,

Similarly for a target language.

$$TL = (\beta_1, \beta_2, \beta_3 \ldots\ldots\ldots \beta_n)$$

Generation of sentences in a natural language say source or target language can be mathematically represented as follows:

$$^{S}{}_{SL} = (S1, S2, S3\ldots\ldots Sn)$$

$$^{T}{}_{TL} = (T1, T2, T3\ldots\ldots Tn)$$

Where $n \rightarrow 1, \infty$

Where $^{S}{}_{SL}$ and $^{T}{}_{TL}$ represents the source and Target language sentences. Hence the relation is

$$^{S}{}_{SL} \in SL \quad \text{and} \quad ^{T}{}_{TL} \in TL$$

When the syntactic rules of a language which generates the sentences can be denoted by G the grammatical and functional relation can be formulated as,

$$G_S : S_{SL} \in SL \quad \text{and} \quad G_T : T_{TL} \in TL$$

Also to generate syntactically correct sentences, a set of pre-defined finite rules called as syntactic rules of a language should operate on symbols or signatures. Each symbol or signature in it embodies a specific meaning and the operation of syntax rules on it completes the generation of a set of meaningful sentences. When each of the symbols called lexical items is subjected to parsing and synthesis we get the stem or basic elements, prefix and suffixes, case endings and inflexions etc.

The Tamil University Machine Translation System (TUMTS) translates technical and scientific texts from a source language of Russian to a target language of Tamil . The functional organization, operation, programming technique involved, implementation and functional features of the TUMTS system is discussed in the following para graphs.

The TUMTS system is designed on the concept of 'TOP DOWN' design approach and it can be modularised as follows. A detailed system functional diagram is shown in figure-4 at the end of this paper.

3.2 COMPUTING RESOURCE :

The Tamil University Computer Centre is equipped with a fourth generation Microprocessor based Micro Computer system. The available Computer system is very small in configuration with a little amount of main memory and auxiliary storage compared with mainframes having 1 or 2 Mega bytes of main memory. More amount of main memory is required for a large volume of dictionary storage and processing. Program transportability to the same group of Computers, software enhancement and software efficiency are considered to be an additional advantage of this system. The Micro-Computer on which the TUMTS system is implemented is an indigenous make having the following configuration.

A)- Hardware

a) 8 bit CPU with Z-80 Microprocessor having multiprocessing facility.

b) 64 Kbytes of main memory per user station.

c) Dual Floppy drives of storage capacity of about 1000,000 bytes (Characters).

d) A dot matrix printer with a speed of about 150 CPS.

B)- Software

a) Multi user multi tasking operating system.

b) BASIC, COBOL, FORTRAN, PASCAL, PL/1 Compilers.

c) System utilities.

BASIC language interpreter has been chosen for coding the TUMTS system. The reason being, it is an interactive language with facilities for debugging and initial program development. As the available main memory for the user program is only 22 Kbytes, the program has to be suitably modularised for providing sufficient space in the memory for operating the source and Target language dictionaries and minor data bases such as Identifier, inflexion and suffix tables.

The dictionary of both source and target language along with identifier, inflexion and case endings table should be available during the translation process.

The system could be able to operate with 1000 vocabularies (Source and target language lexical items, inflexions, case endings, identifier, tables etc). Hence any source language text with a basic vocabulary well below the range of 400 to 500 could be translated by conveniently referring the dictionaries.

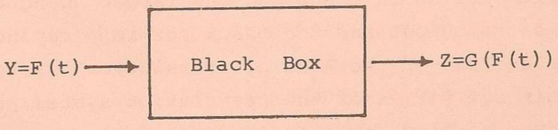
If the operating main memory is available in the range of 512 Kbytes to 1000 Kbytes, it is possible to operate relatively a big dictionary with a large amount of vocabularies.

Therefore this system is designed in such a way that separate dictionaries can be organised and referred for translating the texts belonging to fields such as Physics, Mathematics, Chemistry or General science, Astronomy etc. In other words we can say that the TUMTS system is programmed to access product based dictionaries.

3.3 SYSTEM MODELLING

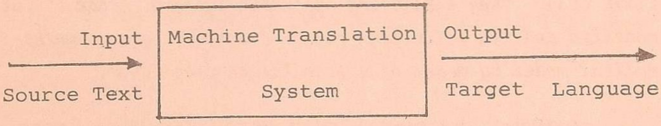
Computer system modelling and simulation are the two most efficient tools available to study and evaluate the intricacies of a system. What types of system can be modelled and how their functions are defined? The definition goes like this *"Any finite and definable system can be modelled and its activities can be described by a mathematical model by means of a formulation procedure"*.

A black box model is a very suitable representation for any of the dynamic system activities.



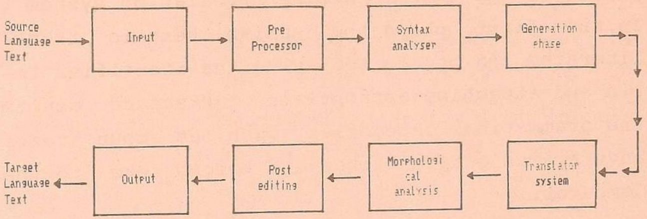
In the above model we find that the input 'Y' is a function of time 't' and output is G(F(t)) i.e. the output is again a function of time. The variable factor may be time, voltage, pressure, velocity etc. In the above representation of a dynamic system the input signal is subjected to synthesis and processing according to the nature and the activities of the system.

So it is obvious that a Machine Translation system is analogous to the above model. The Machine Translation system is represented using the black box technique as follows,



In the above block diagram the source language text acts as an input and the Target Language text is the output and the black box is replaced by the nucleus of the Machine Translation system. The main activities of the translation system are scanning, parsing, analysis and validation of input text, transfer, generation and translation of the source language text, various tables and linguistic attributes supports the translation system to perform the above activities.

The functional organization of a Russian to Tamil Machine Translation System is pictorially represented below:



3.4 PRE - PROCESSOR SYSTEM

For a single lexical item in any of the natural language we find that many of the linguistic attributes are existing. Linguistic data with its grammatical attributes properly defined and structured in a proper syntax contributes a semantically correct sentence. When we are familiar with a particular language it is possible to identify the grammatical attributes such as verb, noun, tense, adverb, adjective, voice, animate or inanimate nouns while scanning the text, but when the same text is analysed by a machine, it is essential to notify the various grammatical patterns of the input text to the Computer.

This procedure is done in two different phases of the TUMTS Machine translation system . In the first phase the Russian text is transliterated and pre-edited using the identifier table and attaching appropriate markers to denote the grammatical categories such as noun, verb, adverb, adjective, tense and voices etc to the Computer.

Thus each of the basic lexical entity is recognised for its syntax with its attributes attached by means of a tag called an identifier. The tag will act as a unique identity for each of the lexical element and thus facilitating the Computer for word and phrase level analysis and translation.

Marking of attributes of lexical item is manually done while transliterating the Russian text. The following identifier table explains the functions of each of the identifiers. They are used for manual editing of the input Russian text.

IDENTIFIER TABLE :

Identifier	Function
#	Representing a nonverbal sentence.

$	Preceding the noun to represent an animate object.
<	To identify the position of the word in the target language.
+	Weak verb identification in relation with Tamil.
@ or 0	Denotes the verbal phrase in Russian.
&	Verbal marker for participle.
X	Number marker for plural.
*	Acts as a delimiter in between verb and adverb or noun and adjective.
*	Passive voice marker in a verbal phrase.

The following Example will illustrate how the various identifiers are used while pre-editing the Russian text.

i) #SOLN*SISTEMA0L DEVJATJ00 PLANETAX0

ii) OCHENJ00<NIZK*TEMPERATURA00<ATMOSFERAXG
GIGANTY*PLANETAXL @POKAZYVAET3 CHTO100
GIGANTY*PLANETAX0 0POLUCHAJUT2
TEPLO00< SOLNCE0R

3.4.1 INPUT TEXT ENTRY :

The Russian text is not of Roman alphabet, hence while transliterating the Russian text using the Roman alphabet it is necessary to pre-edit and mark the grammatical categories. The text can be typed to the Computer one sentence at a time or a number of paragraphs. Text entry is like a normal typing procedure with the help of a video display terminal and a keyboard. The paragraphs are delimited by an identifier '$'.

3.4.2 SOURCE TEXT EDITING MODULE :

Basically the system is designed in a modular way so that each of the module in independent of its activity with the other module and also the entire system is menu driven in nature. This facilitates the user for interactively correcting or editing of the already stored Russian text. The source text Editor module of the TUMTS system performs this activity. It is also possible to edit a particular sentence in the already typed source text by specifying the serial number of the sentence.

3.5 PRE-PROCESSOR MODULE :

Based on the surface structure concept there is a reference structre of the input text and hence all the input source texts should be restructured before the actual translation process. The main objective of the pre-processor system can be listed as follows:

a) To scan and word splitting of the input text.

b) To recognize the syntactic attributes of the sentence.

c) To follow the various rules for analysing the lexical item.

d) To convert the natural form of the input Russian text to the reference structure i.e. from source language to Meta Language form-1.

e) To transform structurally the input Russian text to the pre-specified structure.

The pre-processor system consists of a scanner, parser and lexical analyser. The functions of a scanner system and parser are explained as below.

3.5.1 SCANNER :

The scanner system of TUMTS performs the function of scanning the input Russian text from left to right of a sentence. It makes use of a single blank in between the two words of a source text sentence as a delimiter. Thus when the input source text is taken one sentence at a time the scanner counts the number of words in the sentence and stores it in a temporary table for latter processing. One important thing to be noted at this point is, that the scanner does not identify any of the grammatical attributes of the linguistic data, during this process, but it merely scans for identifying a word for word splitting of the input Russian text.

The scanner module is invoked in the syntax analysis and translation phase of the system. The scanner system can be better explained with the help of the following example,

EX: MARS00 0IZUCHIS&+ BOLJSHE00 LJUDI0I CHEM00
DRUG*PLANETAXA.

When the above sentence is analysed by the scanner system the 'b̸' is treated as a separator of two words and the number of words in the given sentence is 6.

This part of the module helps the pre-processor for Editing the input text. The parsing plays an important role in the pre-processing system.

3.5.2 PARSER :

The parser identifies and recognizes the input text for its grammatical attributes. It is important to attach suitable markers with a word to identify a particular linguistic element for its proper syntax. As explained earlier each of the markers has a specific function to do when they are associated with a lexical element. With the help of the identifiers the parser will be able to recognize the lexical item.

For example let us consider a sentence below

ZEMLJA0R 0HOROSHO*VIDIAT6*MOREX0 TIP00

TJOMN*OBLASKOX0

When the scanner completes its operation of splitting the input sentence into words,the parser module is invoked. In the parser system words are taken one at a time and are recognized with the help of identifiers attached with the word. In the above sentence '0' in 0HOROSHO*VIDIAT6*represents verb phrase and a 'X' at the end denotes that it is a passive voice. During the course of recognition of each of the words, the parser makes use of the identifier table and a set of syntactic rules.

Once the parser system recognizes the basic lexical element it is always possible to classify the linguistic data into verb, noun, comparative phrase, tenses etc. So the pre-editing part of the pre-processor recognizes, locates and restructures the original Russian input text i.e converting to an intermediate form called as Meta language-1.

By means of a structural transformation procedure using the rules of the Meta language-1 the syntactic construction is derived in this part of the system. Here the pre-edited text, irrespective of the natural language form with its verb

always appearing as a second unit is more suitable for acting as an input to the Translation phase of the system. That means the translation module of the TUMTS system can accept the input text having a standard syntax only. Therefore the function of the pre-processor is analysing, identifying and restructuring of a given input Russian sentence to a standard syntactic format.

Before actually restructuring can take place on the given input Russian sentence the pre-editor processes the following forms of sentences: compound sentences, sentences with comparative phrases and sentences with genitive cases etc. In the first phase it restructures the input text based on the above syntactic categories and in the second phase the pre-processor relocates the verbal and noun phrases so that a standard syntactic structure is formed. The activities can be explained in the following examples.

EX : (i) with genitive case :

Input
source
text : RAZMERA00<MASSA00<PLOTNOSTJ00<VENERA0G

@BLIZKJUT2 ZEMN*TIP00 PLANETAXA

Pre-processed text (Meta language-1)	VENERA0G @BLIZKJUT2 RAZMERA00 MASSA00 PLOTNOSTJ00 ZEMN*TIP00 PLANETAXA

(ii) with comparative phrase:

Input source text	MARS00 0IZUCHIS&+ BOLJSHE00 LJUDI0I CHEM00 DRUG*PLANETAXA

Pre-processed text (Meta language-1)	MARS00 0IZUCHIS&+ DRUG*PLANETAX0 CHEM00 BOLJSHE00 LJUDI0I

3.6 LANGUAGE DATA BASE :

In a Machine Translation system the major constituents are Dictionaries for source and target languages and a set of rules and minor data bases for the Computer to understand the syntactic structure of both the languages. Besides this a set of algorithmic rules and definitions will have to be provided for interpreting, transforming, analysing, generating and translating the source text to the target language.

The major language data bases are

i) Source language dictionary

ii) Target language dictionary

and minor data bases are :

i) Identifier table

ii) Inflexion table

iii) Verbal suffix table

In the TUMTS system language data base, not much of the grammatical information in the form of grammar codes are provided in the dictionary, instead much of the grammatical definitions are provided in algorithmic part of the system. Some of the grammatical attributes are available in the minor data bases such as identifier table and inflexion tables. By means of a table look up procedure using minor data bases the translation and generation phase of the system recognizes the basic grammatical attributes.

As the TUMTS system is meant for technical literature translation, it is advised to organize a product based dictionary. A suitable schema is provided for easy access, updating and addition and easy look up while translation is in progress.

The functions of the language data bases are:

i) The major data bases provide equivalent translation for word to word.

ii) Some of the grammatical codes, attached with the linguistic data aids in identification and analysis during the translation procedure.

iii) It facilitates validating the lexical element.

iv) The minor data bases help in validating the lexical element and also aid in the generation phase.

v) Identifier table acts as a major look up list for syntactic recognition during the parsing.

vi) Inflexion table provides equivalent case endings in the target language.

3.6.1 STRUCTURE OF THE DATA BASES

Basically the TUMTS system consults a bilingual Dictionary of Russian to Tamil during

the translation process. The major data bases for Russian and Tamil are organised in the following manner.

i) Only stem of the verb for source & target language is stored in the dictionary.

ii) Some of the minor grammatical identifiers for animate or inanimate nouns are provided in the dictionary.

iii) Only singular forms of nouns are stored in the dictionary.

iv) Identifier table contains mostly special symbols with its functional attributes defined in the main system.

v) Case endings and suffixes in Russian are coded for simplification and resolving the semantic ambiquities.

vi) The dictionary is organized in a classified way with separate tables for nouns, verbs, adverbs, adjectives and case endings etc and their locations are properly indexed for easy access during translation.

A listing of sample dictionary is attached in appendix for reference. Due to the limitation of the main memory in the available Computer system the present TUMTS system can handle about 1000 vocabularies all in total. As this system has a product based dictionary, it is suggested to organize different data bases for different fields of application. For example when the system translates physics articles from Russian to Tamil a separate bilingual dictionary may be consulted and similarly for Engineering, the system may be instructed to refer to appropriate dictionary.

3.7 TRANSLATION PHASE :

The most complicated part of the Machine Translation system involves a translator and generator module. A translator module performs functions such as scanning, parsing, morphological analysis, table look up and finally translation etc.

The output of the pre-processor module will be the input of the translator module The translator module operates only on an input Russian text with a standard syntax otherwise it

will not be able to recognize and process the text. In other words we can say that the translation phase will process only an intermediate language. Here the intermediate language is nothing but the Meta language form-1. The activities of the translator system can be summarized as below:

i) Scanning and parsing the pre-processed text and carrying out lexical analysis.

ii) Table look up for data validation and word synthesis on them.

iii) Identifying and processing of verbal and non verbal phrases.

iv) Segmenting the word into stem, inflexion, suffix and identifier and carrying out morphological analysis.

v) By referring the suffix and case endings tables the equivalent forms are chosen by the Translator from the Target language tables.

vi) Displaying error messages in case of word not available in the dictionary after validation process.

vii) Resolving some of the semantic ambiquities.

viii) Correcting the spelling of input text(Meta language form-1) while translation is in progress.

ix) Word to word translation after validating the data.

x) Translating proper names and numbers.

xi) Transfer of intermediate output to the generation phase.

The main function of the Translator **module** can be explained as follows. The pre-processed text is available in a file with delimiters for sentence, paragraphs as well as the number of units in a sentence. The input of the Translator is the Meta language-1 text and the translator takes one sentence at a time for processing. The scanner system scans the Russian text from left to right and the task of the parser is to identi fy and recognize each of the phrases in the input sentence.

If a '#' sign preceds the sentence then the syntactic analyser recognizes it as a

nonverbal sentence, otherwise the verbal sentence is passed on for further processing to the verbal processor module. The basic functional difference between the varbal and nonverbal sentence processing is that the verbal sentence includes synthesis and validation, transfer and generation of equivalent verbal phrase in Tamil. If an adverb is also preceding a verb with a delimiter of '*' then it will be isolated and equivalent in Target language is chosen before the verb is taken for processing.

3.7.1 VERB PHRASE ANALYSER :

Lexical analyser identifies the grammatical attributes of the element with the help of an identifier and after synthesis of the verb other grammatical attributes of suffixes for tense voice and participles are morphologically analysed. A given verb phrase in Russian is synthesised into stem, tense marker, voice marker and suffixes. The equivalents in Tamil are searched in the target language dictionary. After the processing is over this intermediate translated version in Tamil is transferred to the generation phase. The following example will explain the verb phrase analysis of a translator module.

EX: Verbal phrase

```
0HOROSHO*VIDIAT6*
 ------- ----  ---
```

i.e 0 is the verb phrase identifier and before processing the verb, the adverb HOROSHO will be eliminated from the phrase and its equivalents will be searched in the Target language dictionary. The equivalent result is stored in a temporary location of the memory. The adverb and verb are isolated by means of a '*' in between them. Then the verb phrase VIDIAT6* is synthesised for morphological analysis into the main part VIDI as stem and AT6 as the suffix and '*' as the passive voice marker. Then the equivalents are chosen from the Target language dictionary for VIDI and AT6. Thus for a given verb phrase the analysis and translation occurs as follows :

0HOROSHO*VIDIAT6* ==> Synthesis & Morphological Analysis.

Translation--->

i.e

HOROSHO ==>NANRAAKA ; VIDI ==>PAAR ; AT6 ==> KINRANA

3.7.2 NON VERB PHRASE ANALYSER :

This part of the TUMTS translator system involves lexical analysis, word synthesis, morphological analysis, table look up procedure for data validation and transfer

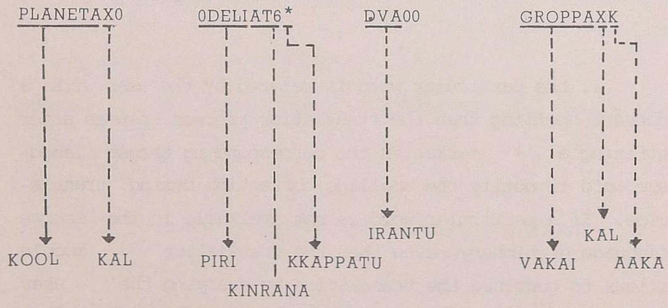
etc. During the lexical analysis a particular phrase is analysed and any associating grammatical category like an adjective is identified. If there is an adjective preceding a noun phrase with a delimiter of '*' then the equivalents in target language is obtained after validating the data and by referring to the language data bases.

After the lexical analysis the noun phrase is synthesised for stem, case marker and number marker etc. The stem of the noun in the source language is validated by referring to the source language dictionary. If there is any spelling error, then the system will interactively query the user for correcting the spelling by on line.

If the particular word is retyped by the user with a correct spelling then the translation process resumes after affixing a '+' marker in the corresponding target language word to notify the spelling correction during translation. If a particular word is not available in the source language dictionary, even then the translation module allows to continue the translation by querying the user interactively to type the Target equivalents of the word. This process enables the system to create a temporary dictionary for reference and translation. The above process occurs interactively by displaying suitable messages for the user to intervene and help the translation process.

Once if the stem of the noun is validated then the equivalent is obtained from the target language dictionary and their corresponding case ending is referred from the target inflexion table. The plural or singular value is attached with the translated Tamil equivalents by the algorithmic processor. For example consider the following sentence, and see how the translated version of the output will look, after the above phase of operation is over.

EX : Input text in Meta language form - 2



The various constituents of the Translating phase can be briefly explained as follows :

3.7.3 SYNTAX ANALYSER :

It has both scanner and parser system to scan the source text sentence and to recognize the source text for its structural syntax. It makes use of Identifier table, source and target language data bases during this process. Syntax analyser module identifies the major sentence structure and also classifies the text into verbal and nonverbal sentences. During this process it also carries out the lexical analysis.

3.7.4 LEXICAL ANALYSER :

In the translation phase the function of a lexical analyser is to identify and analyse the individual grammatical categories of the phrases irrespective of the major syntax. It uses major data bases and inflexion and identifier tables. Lexical analysis helps in synthesising the phrase before translation.

3.7.5 MORPHOLOGICAL ANALYSER :

It is the linguistic procedure by which a given word during the process of translation is subjected to synthesis and grammatical analysis. Thus when a given phrase is synthesised into stem, suffix, inflexion etc, they are lexically identified and analysed during the morphological analysis .

3.7.6 RESOLVING SEMANTIC AMBIQUITIES :

In the present system the semantic ambiquities never arise since they are properly codified in the intermediate language itself. Hence, based on the suffix value the translator will choose the correct equivalents of target language for a given source word.

3.7.7 WORD TO WORD TRANSLATION :

It is only a simple table look up procedure. The input source language text in the Meta language form-1 is synthesised and subjected to lexical and morphological analysis and the basic element is converted to a stem and inflexion. By using the stem it is possible to search for a match in the source language dictionary. It is the procedure of validating the data. The source and target language dictionaries contain only the root of the noun and verb and the infinitives are generated by the system with the help of the stem and a set of rules. Since the validation is being a search and match process, the translation module is programmed to display suitable messages for human interaction in case if it finds a spelling error or nonmatching of the stem with the entries of the source language dictionary.

Further, word to word translation is facilitated by providing suitable dictionary reference indexes in the

Restructured output ⟹ structural transfer for proper syntax.

Output :

VAAYU ATUKKUTAIYA AALAM PERIYA KOOLKALIL MIKA ATIKAMAAYIRUKKIRATU.

3.9 POST EDITING :

Due to the grammatical pecularities of the Tamil i.e inflexions and 'Santhi' the translated output text will not be very close to the natural language text. Hence it necessiates Post Editing of the output target language text. This may be done manually also, but to some extent it is possible to do post editing with the help of a Computer. In the present system there is no separate module for post editing. Instead the post editing is done simultaneously during the Generation phase. Due to the limitation of the main memory and the stylistic variation of the input Russian text, the present system is not programmed to handle the post editing.

Post editing is automatized for Tamil words with 'makara' ending. The machine will post edit the text only in conjunction with the case ending and the number marker. Post editing takes place when the logical AND ing of case ending and singular markers is true, otherwise no post editing will occur.

For example :

Before Post editing :

CUURIYA MANTALAMIL ONPATU KOOLKAL

Here it is found that the logical AND ing of case ending i.e IL and singular marker is true. Hence the post edited text will be as follows :

CUURIYA MANTALATTIL ONPATU KOOLKAL - Case ending with singular.

3.10 OUTPUT PHASE :

The output can be either viewed on a vedio display terminal or printed out as a hard copy in a printer. While translation is one sentence at a time or a group of paragraphs, the translated text in the target language will be stored in the disc. The output in Tamil if displayed on the screen or printed on the printer will be in Roman script. When a Tamil based script facility is available in the Computer, it is possible to have the translated text in Tamil script printed on a paper in more natural form.

3.11 CASE STUDY :

In order to test the operating efficiency of the TUMTS system, one page of technical text in Russian consisting of 45 sentences was transliterated, pre-edited and fed

as an input to the TUMTS program. The input text had excellantly gone through pre-processing, translation , generation and post editing phases of the TUMTS system.

3.11.1 EXPERIMENTAL OBSERVATION :

An one page translated Tamil text in Roman script was the result of the above process. The original Russian text, transliterated and pre-edited text and the translated Tamil text are enclosed for reference in the appendix. To translate the 45 Russian sentences with 227 words the TUMTS system took approximately 4 minutes and 30 seconds. The operating speed of the system is not a linear function of the volume of input text. The speed of the TUMTS system is a function of the input text and the main memory.

During the course of this system development project particularly at the beginning some of the interesting features were observed. With 25 nouns and 9 pronouns the machine was programmed with certain Russian syntax rules for automatic generation of sentences. The nouns included singular, plural, all the genders and animate and inanimate. The Computer generated about 225 sentences. But not all the sentences are syntactically correct. So by elimination procedure further rules were programmed to eliminate syntactically incorrect sentences. Finally the machine generated 77 syntactically and semantically acceptable sentences.

During the testing it was observed that operating on a fixed memory working space, longer sentences with more strings took relatively large amount of time for translation. If more memory is available for workspace then the translation is faster.

The TUMTS system is programmed to process the following attributes of the input Russian text.

a) A sentence with a maximum of about 16 units.

b) Passive voice.

c) Present/past/past participles.

d) All the cases.

e) Numeral(s) and proper names.

f) Plurals.

g) Generating verb and Noun phrases from the stem available in the dictionary.

h) Automatic pre-editing for structural transfer of input text to the Meta Language form-1

i) Able to handle compound sentences.

j) Can process comparative and genitive phrases.

k) Post editing facility is provided.

algorithmic definition for various grammatical elements such as noun, verb, adverb and adjectives.

If the validation is successful after word matching, then the equivalents in Tamil is chosen and stored in a temporary table for latter processing. Hence the word to word translation is merely a substitution procedure.

3.8 GENERATION PHASE :

The main function of the Generation phase is to generate the word, phrases and sentences in the target language. During this phase, the activities of generation module are , referring to the source and target language syntax and the structural transformation etc. Generation phase of the TUMTS system can be sub-modularised into generation module, Interface module, source and target language syntax and structural transfer module.

3.8.1 GENERATION MODULE :

In the Translation phase of the system word to word translation of the Input source language text takes place so also for the identifiers, inflexions and case endings the equivalents are generated using the major target language data base and other minor data bases. The intermediate form of the target language sentence will not be in a proper syntax at this stage. The string phrases so translated are concatenated using the syntax rules of the target

language. Thus the generation phase transforms and generates semantically and syntactically acceptable phrases of the Target language. It is necessary that each of the phrase generated by this module should refer to the lexical rules of the target language. Considering the same example, it is found that the generated target language output will be of the form:

EX : Pre-edited Input text : Meta language form-1

PLANETAX0 ODELIAT6* DVA00 GRUPPAXK

Generated text :

KOOLKAL PIRIKKAPPATUKINRANA IRANTU VAKAIKALAAKA.

3.8.2 STRUCTURAL TRANSFER MODULE :

Though the translation and generation phase synthesises the input Russian text and the resultant output will be the word to word translation of the actual pre-edited input text, the output will be in an intermediate language called Meta language-2. So the phrases of the Meta language-2 should have to be suitably located to form a syntactically and semantically correct sentences. It is emphasised that surface structure theory of the target language should be adopted for suitably locating the target language phrases. To explain how a structural transfer occurs in the Meta language-1 text to get the final output,

let us consider the pre-edited input text or Meta language-1 text represented as,

S = (S1, S2, S3, S4, S5, S6, S7)

Where S1, S2 .... S7 represents the phrases in the pre-edited input text and S2 ---> verb phrase.

After translation and generation, the target language text in the Meta language form-2 is represented as,

T = (T1, T2, T3, T4, T5, T6, T7)

Where T1, T2, T3, T4, T5, T6, T7 are target language phrases with the sentence having an intermediate target structure. In the surface level the translation is simply word to word i.e having one to one correspondance.

<u>Translation</u>

S1 ---------------> T1
S2 ---------------> T2
. .
. .
S7 ---------------> T7

After effecting the structural transfer the previously shown translated sentence will be restructured as

KOOLKAL IRANTU VAKAIKALAAKA PIRIKKAPPATU KINRANA.

Therefore the syntactically correct version of the translated sentence in Tamil will be represented by T as follows:

T = (T1, T3, T4, T5, T6, T7, T2)

Where T2 ---> verb phrase in Tamil.

Here T1, T3, T4, T5, T6, T7, T2 represents the surface structure of the target language.

It will be more appropriate if we explain the entire operation of the translation system such as pre-processing, translation, Generation and structural transfer using the following example :

Input : GLUBINA0<GAZ0V*SLOJ0G GIGANTY*PLANETAXL
OCHENJ00 @VELIKET3

Meta lang-1 => Pre-processing

SLOJ0G @VELIKET3 GLUBINA00 GAZ0V

GIGANTY*PLANETAXL OCHENJ00

Intermediate form of output in Tamil = = = =>
Translation & Generation

Meta Language-2

VAAYU ATIKAMAAYIRUKKIRATU ATUKKUTAIYA AALAM

PERIYA KOOLKALIL MIKA

This system is programmed to solve only certain grammatical attibutes of the Russian Language text. To develop a system that can work for all the grammatical pattern in Russian language, the rules to be provided must be very exhaustive. Due to the Limited resources available and with a limited main memory it was found very difficult to include all the natural inflexions of the Russian language.

Further it was found that if the dictionary is added with even a smaller number of vocabularies the system becomes very slow and even at sometimes the Computer crashes due to the non availability of main memory for the working space.

The infinitives and future tenses are not handled in the present system. It is also proposed to refine and improve the system for adding new grammatical rules.

3.11.2 PROGRAMMING MODULES & FILES :

As explained earlier the entire software system is designed in a more modular way. The present software system includes three programming modules for pre-processing text editing and translation.

It makes use of data files for input text entry, pre-processed text and the target language output. Apart from this it makes use of data files for major and minor data bases to store and access the bilingual information. Most of the data files are Random access files for easy and fast access during the translation phase of the system.

The following table illustrates the various programming and data files with its functional definition.

File Label	Functional definition
Programming files	
i) EDITOR.BAS	To pre-edit input Russian text.
ii) TRANSLAT.BAS	This is the heart of the TUMTS system. It translates the Russian text into Tamil when it is invoked.
iii) OUTPUT.BAS	To display the output on a screen as well as to print the output. Also this module edits the input source text.

Data files :

iv)	SOURCE1.DAT	It stores the Edited input Russian text.
v)	SOURCE2.DAT	This file contains the output of the pre-processor.
vi)	TARGET.DAT	After effecting the translation system the output target language text is stored in this file.
vii)	DICTION1.DAT	It is a bilingual Russian-Tamil Dictionary containing verb, Noun and their Indexes for access.
viii)	DICTION2.DAT	This is again a bilingual Dictionary containing inflexions, suffixes and identifiers.

In the beginning of the system development project the TUMTS system was able to perform the translation from Russian to Tamil and vice versa, for simple sentences of the form NP+NP & NP+VP. But when it was tried to transform from a simple source language surface structure to the semantic deep structure, it was not possible to program for all the syntax rules. Hence it was decided to concentrate on programming the system for different structure of the input source text of a Russian to Tamil Translation system.

3.12 CONCLUSION :

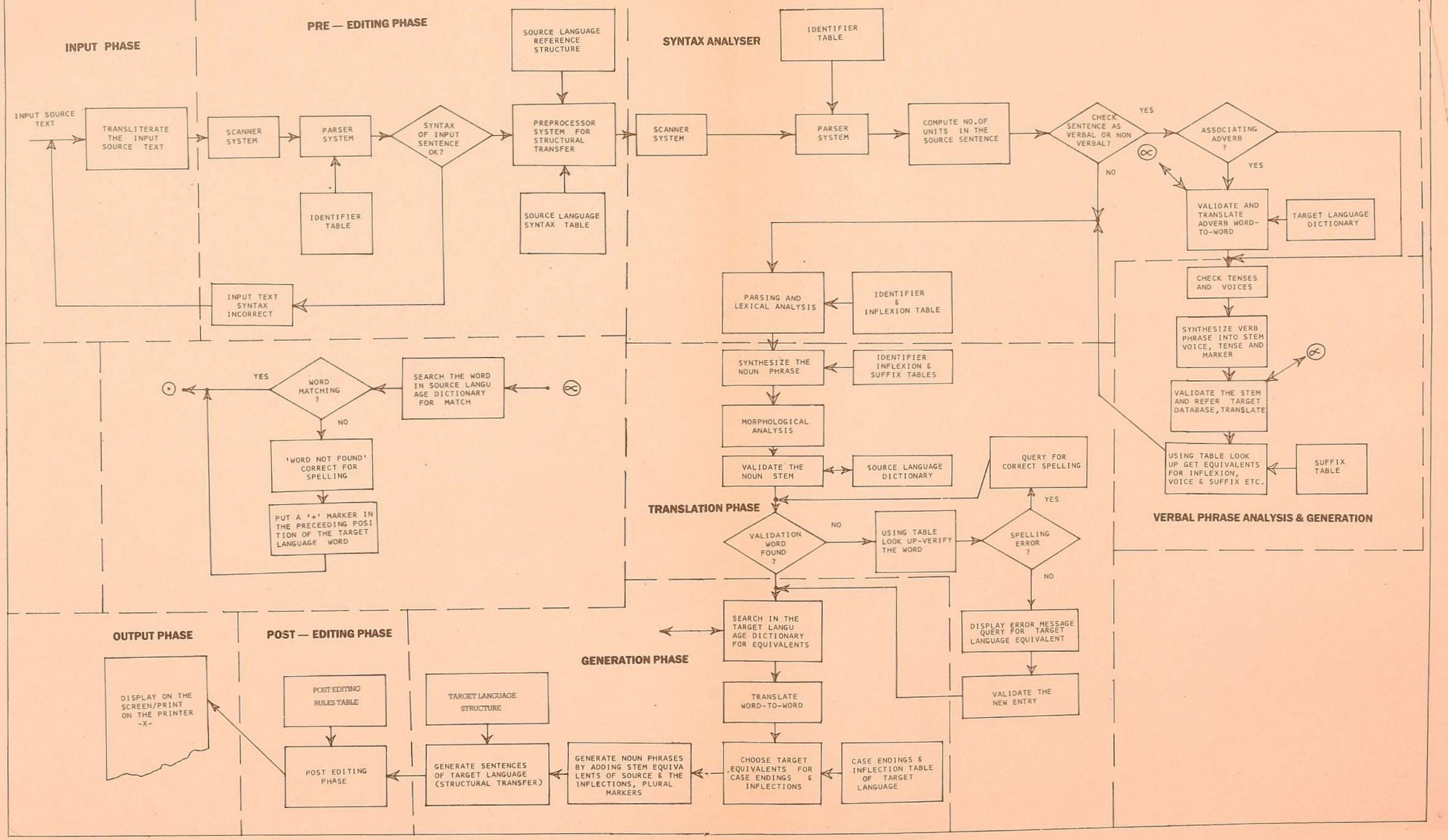
Having programmed for certain rules of the Russian and Tamil languages of the TUMTS translation system, it is also possible to codify the existing system with minimum effort for Tamil to Russian translation. This can be achieved in the present system if the main memory of the Computer is enhanced.

APPENDIX — A

TAMIL UNIVERSITY MACHINE TRANSLATION SYSTEM

RUSSIAN TO TAMIL

FUNCTIONAL FLOW DIAGRAM



APPENDIX B.

###

BILINGUAL DICTIONARY RUSSIAN - TAMIL

###

*** NOUNS ***

A	AANAAL
AMMIAK	AMMOONIYAA
ATMOSFERA	KAARRUMANTALAM
BLIZHE	ARUKIL
BOLJSHE	PERITUM
CHEM	KAATTILUM
CHETYRE	NAANKU
CHTO1	ENRU
CVET	NIRAM
DALEKO	TUURATTIL
DALJSHE	TUURATTIL
DAVLENIE	ALUTTAM
DENJ	PAKAL
DEVJATJ	ONPATU
DIAMETR	VITTAM
DLINNEE	ATIKANIILAM
DVA	IRANTU
EGO	ATANUTAIYA
ESHCHJO	INNUM
GALILEJ	KALILIYOO
GAZ	VAAYU
GELIJ	HIILIYAM
GIGANTY	MIKAPPERIYA
GLUBINA	AALAM
GOD	AANTU
GRUPPA	VAKAI
HOLODNEE	ATIKAKULIR
JUPITER	VIYAALAN
KANAL	KAALVAAY
KLIMAT	TATPAVEPPAM
KM	KI.MII
KOLICHESTVO	ALAVU
LETO	KOOTAIKKAALAM
LINIJA	VARI
LJUDI	MANITARKAL
LOMONOSOV	LAMANOSOV
LUNA	CANTIRAN
MARS	CEVVAAY
MASSA	ETAI
MERKURIJ	PUTAN
METANA	METANAAL
MORE	KATAL
NEMNOGO	KONCAM
NEPTUN	NEPTUUN
NET	ILLAI
OBLAKO	MEEKAM
OBLASTJ	PIRATEECAM

OCHENJ	MIKA
OSENJ	ILAIYUTIRKAALAM
OSJ	ACCU
PLANETA	KOOL
PLOTNOSTJ	ATARVU
PLUTON	PULUUTOON
POLJUS	TURUVAM
POVERHNOSTJ	MEELPARAPPU
RAZMERA	PARIMAANAM
SATURN	CANI
SHAPKA	MUKAL
SISTEMA	MANTALAM
SLOJ	ATUKKU
SOLNCE	CUURIYAN
SPUTNIK	TUNAIKKOOL
SVOJ	TANNUTAIYA
TEMPERATURA	VEPPANILAI
TEPLO	VEPPAM
TIP	POONRA
TOLJKO	TAAN
UCHONIJ	VINGAANI
UGLEKISL	KARIYAMILA
URAN	UURANAS
VENERA	VELLI
VESNA	VACANTAKAALAM
VODOROD	NIIRVAAYU
VOKRUG	CURRI
VSE	ELLAA
ZEMLJA	PUUMI
ZIMA	KULIRKAALAM

*** VERBS ***

BLIZK	OTTIRU
CHETA	PATI
DELA	CEY
DELI	PIRI
DUMA	NINAI
ES	IRU
IGRA	VILAYAATU
ISPYTYVA	PERU
IZMENIA1	MAARRU
IZMENIA2	MAARU
IZUCHI	AARAAYA
NABLUDA	URRUNOOKKU
NAHODI	KAAN
NAHODI1	TEETU
NAHODI2	IRU
NAJTI	KANTUPITI
OBNARUZHI	KANTUPITI

OKRUZHI	CUULA
OTDYHA	OOYVETU
OTKRY	KANTUPITI
OTVECHA	PATILCOL
POKAZYVA	KAATTU
POKUPA	VAANKU
POLUCHA	PERU
RABOTA	VEELAICEY
SLUSHA	KEET
SOSTOI	IRU
SPRASHIVA	KEET
STROI	KATTU
VELIK	ATIKAMAAYIRU
VIDI	KAAN
VIDI	PAAR
VRASHCHA1	CULARRU
VRASHCHA2	CULAL
ZAKRY	MOOTU

*** ADJECTIVES ***

BEL	VELLAI
BLIZK	OTTA
BOLJSH	ATIKA
CHEM	KAATTILUM
DEVIAT	ONPATU
DRUG	MARRA
FIOLETOV	UUTAA
GAZOV	VAAYU
GIGANTY	PERIYA
GLAVN	MUKKIYA
HIMICHESK	VEETIYAL
HLORIST	KULOORAIT
INTERESN	CUVAIYAANA
JARK	PRAKAACAMAANA
JARKOGOLUB	PRAKAACAMAANA-NEELAM
KRASN	CIVAPPU
NEKOTOR	CILA
NIZK	KURAINTA
NOVIJ	PUTIYA
ODINAKOV	OREEMAATIRIYAANA
OGROMN	ATIKA
PERV	MUTAL
RAZLICHN	VEERUPATTA
RAZLICHN	VEVEERU
SOLN	CUURIYA
SPEKTRALJ	NIRAMAALAI
TAK	IPPATI
TJOMN	IRUNTA
UGLEKISL	KARIYAMILA

VELIK	ATIKAM
VERHN	VELI
VNUTREN	UL
VTOR	IRANTAAVATU
ZAMECHATELJ	ARPUTAMAANA
ZELJON	PACCAI
ZEMN	PUUMI
ZHJOLT	MANCAL

*** ADVERBS ***

CHASTO	ATIKKATI
DAVNO	MUNPEE
GROMKO	CATTAMPOOTTU
HOROSHO	NANRAAKA
MALO	KURAIVAAKA
MEDLENNO	METUVAAKA
NEMEDLENNO	UTANEE
POSTOJANNO	TOTARNTU
PRAVILNO	CARIYAAKA
PRAVILNO	CARI
PRJAMO	NEERAAKA
TIHO	AMAITIYAAKA
TRUDNO	KATINAM
TVJORDO	TELIVAAKA

*** PRONOUNS ***

ETO1	ITU
ETO2	IVAN
ETO3	IVAL
ETO4	IVARKAL
ETO5	IVAI
JA	NAAN
MY	NAANKAL
ON1	IVAN
ON2	ITU
ONA1	IVAL
ONA2	ITU
ONI1	IVARKAL
ONI2	IVAI
ONO	ITU
TY	NII
VY	NIINKAL

*** SUFFIXES ***

EM	KIROOM
ESH	KIRAAI
ET1	KIRAAN
ET2	KIRAAL
ET3	KIRATU
ET4	KIRAAR
ETE	KIREERKAL
JU	KIREEN
JUT1	KIRAARKAL

JUT2	KINRANA
L1	TTAAN
L2	TTAAR
L3	TTEEN
L4	TTAAI
LA	TTAAL
LI1	TTAARKAL
LI2	TTOOM
LI3	TTIIRKAL
LI4	TTANA
LO	TTATU

AT5	KIRAARKAL
AT6	KINRANA
L1	TTAAN
L2	TTAAR
L3	TTEEN
L4	TTAAI
LA	TTAAL
LI1	TTAARKAL
LI2	TTOOM
LI3	TTIIRKAL
LI4	TTANA
LO	TTATU
M1	KIROOM
SH	KIRAAI
T1	KIRAAN
T2	KIRAAL
T3	KIRATU
T4	KIRAAR
TE	KIREERKAL
U1	KIREEN

*** CASE ENDINGS ***

O	
A	AI
D	UKKU
G	UTAIYA
I	AAL
K	AAKA
L	IL
M	ITAM
N	IN
P	PARRI
R	ILIRUNTU
S	UDAN

##

APPENDIX C.

Types of declensions-A comparative analysis

Type 1:

Singular Masculine

Animate

Case	In Russian	In Tamil	Inflexion Russian	Inflexion Tamil
Nominative	SLON	AANAI	Nil	Nil
Genitive	SLONA	AANAIYIN	-A-	-IN-
Dative	SLONU	AANAIKKU	-U-	-KKU-
Accusative	SLONA	AANAIYAI	-A-	-AI-
Instrumental	SLONOM	AANAIYINAAL	-OM-	-INAAL-
Associative	SO SLONOM	AANAIYUTAN	-OM-	-UTAN-

Inanimate

Case	In Russian	In Tamil	Inflexion Russian	Inflexion Tamil
Nominative	STOL	MEEJAI	-	-
Genitive	STOLA	MEEJAIYIN	-A	-IN
Dative	STOLU	MEEJAIKKU	-U	-KKU
Accusative	STOL	MEEJAIYAI	-	-AI
Instrumental	STOLOM	MEEJAIYINAAL	-OM	-INNAL
Associative	SO STOLOM	MEEJAIYUTAN	OM	-UTAN

Neuter

Nominative	SLOVO	VAARTTAI	O	-
Genitive	SLOVA	VAARTTAIYIN	A	IN
Dative	SLOVU	VAARTTAIKKU	U	KKU
Accusative	SLOVO	VAARTTAIYAI	O	AI
Instrumental	SLOVOM	VAARTTAIYINAAL	OM	INAAL
Associative	SLOVOM	VAARTTAIUTAN	OM	UTAN

Plural Masculine-animate

Case	Russian	Tamil	Inflexion	
			Russian	Tamil
Nominative	SLONY	AANAIKAL	Y	-
Genitive	SLONOV	AANAIKALIN	OV	IN
Dative	SLONAM	AANAIKALUKKU	AM	UKKU
Accusative	SLONOV	AANAIKALAI	OV	AI
Instrumental	SLONAMI	AANAIKALINAAL	AMI	AAL
Associative	SO SLONAMI	AANAIKALUTAN	AMI	UTAN

Inanimate

Nominative	STOLY	MEEJAIKAL	Y	-
Genitive	STOLOV	MEEJAIKALIN	OV	IN
Dative	STOLAM	MEEJAIKKU	AM	UKKU
Accusative	STOLY	MEEJAIKALAI	Y	AI
Instrumental	STOLAMI	MEEJAIKALINAAL	AMI	INAAL
Associative	SO STOLAMI	MEEJAIKALUTAN	AMI	UTAN

Plural neuter

Nominative	SLOVA	VAARTTAIKAL	A	
Genitive	SLOV	VAARTTAIKALIN	-	IN
Dative	SLOVAM	VAARTTAIKALUKKU	AM	UKKU
Accusative	SLOVA	VAARTTAIKALAI	A	AI
Instrumental	SLOVAMI	VAARTTAIKALINAAL	AMI	INAAL
Associative	SO SLOVAM	VAARTTAIKALUTAN	AMI	UTAN

Feminine Singular

Animate

Case	Russian	Tamil	Inflexion Russian	Inflexion Tamil
Nominative	ZHENSHCHINA	PEN	A	-
Genitive	ZHENSHCHINY	PENNIN	Y	IN
Dative	ZHENSHCHINE	PENNUKKU	E	UKKU
Accusative	ZHENSHCHINU	PENNAI	U	AI
Instrumental	ZHENSHCHINOJ	PENNAAL	OJ	AAL
Associative	SO ZHENSHCHINOJ	PENNUTAN	OJ	UTAN

Type 2.

Feminine singular inanimate

Nominative	STRANA	NAATU	A	-
Genitive	STRANY	NAATTIN	Y	IN
Dative	STRANE	NAATTUKKU	E	UKKU
Accusative	STRANU	NAATTAI	U	AI
Instrumental	STRANOJ	NAATTINAAL	OJ	AAL
Associative	SO STRANOJ	NATTUTAN	OJ	UTAN

Plural animate

Nominative	ZHENSHCHINY	PENKAL	Y	-
Genitive	ZHENSHCHIN	PENKALIN	-	IN
Dative	ZHENSHCHINAM	PENKALUKKU	AM	UKKU
Accusative	ZHENSHCHIN	PENKALAI	-	AI
Instrumental	ZHENSHCHINAMI	PENKALINAAL	AMI	AAL
Associative	ZHENSHCHINAMI	PENKALUTAN	AMI	UTAN

Plural inanimate

Nominative	STRANY	NAATUKAL	Y	-
Genitive	STRAN	NAATUKALIN	-	IN
Dative	STRANAM	NAATUKALUKKU	AM	UKKU
Accusative	STRANY	NAATUKALAI	Y	AI
Instrumental	STRANAMI	NAATUKALINAAL	AMI	AAL
Associative	STRANAMI	NAATUKALUTAN	AMI	UTAN

Type 3.

Feminine with soft ending singular

Case	Russian	Tamil	Inflexion	
			Russian	Tamil
Nominative	LOSHADJ	KUTIRAI	J	-
Genitive	LOSHADI	KUTIRAIYIN	I	IN
Dative	LOSHADI	KUTIRAIKKU	I	KKU
Accusative	LOSHADJ	KUTIRAIYAI	J	AI
Instrumental	LOSHADJJU	KUTIRAIYAAL	JJU	AAL
Associative	LÓSHADJJU	KUTIRAIYUTAN	JJU	UTAN

Plural

Nominative	LOSHADI	KUTIRAIKAL	I	-
Genitive	LOSHADEJ	KUTIRAIKALIN	EJ	IN
Dative	LOSHADJAM	KUTIRAIKALUKKU	JAM	UKKU
Accusative	LOSHADEJ	KUTIRAIKALAI	EJ	AI
Instrumental	LOSHADJAMI	KUTIRAIKALAAL	JAMI	AAL
Associative	LOSHADJAMI	KUTIRAIKALUTAN	JAMI	UTAN

APPENDIX D.

*** ORIGINAL RUSSIAN SOURCE TEXT ***

Планеты солнечной системы

В солнечной системе девять планет: Меркурий, Венера, Земля, Марс, Юпитер, Сатурн, Уран, Нептун, Плутон. Планеты делятся на две группы: планеты земного типа — Меркурий, Венера, Марс и планеты-гиганты — Юпитер, Сатурн, Уран, Нептун. Плутон был открыт только в 1930 году и ещё очень мало изучен.

Меркурий находится к Солнцу ближе, чем другие планеты. Меркурий немного больше Луны: его диаметр 4840 км, а диаметр Луны — 3476 км. Атмосферы на Меркурии нет.

Размеры, масса и плотность Венеры близки к земным. Венера окружена атмосферой, которая делится облаками на два слоя. В верхнем слое найдено большое количество углекислого газа. Атмосфера этой планеты была открыта в 1761 году М. В. Ломоносовым. Венера медленно вращается вокруг своей оси. День на ней длиннее, чем на Земле.

Марс изучен людьми больше, чем другие планеты. Поверхность Марса постоянно наблюдается учёными. С Земли хорошо видны белые шапки полюсов, тёмные области «морей», линии «каналов». Зимой, весной, летом и осенью цвет поверхности планеты изменяется. Марс находится от Солнца дальше, чем Земля. Климат на нём холоднее, чем на Земле.

Атмосферы обнаружены и на планетах-гигантах. Они состоят из метана, аммиака, водорода, гелия. Глубина газового слоя на планетах-гигантах очень велика. Внутренние слои атмосферы на Юпитере, Сатурне, Уране, Плутоне испытывают огромное давление верхних слоёв газов. Очень низкая температура атмосфер планет-гигантов показывает, что они получают тепло от Солнца, от которого они очень далеко находятся.

Поверхность этих планет закрыта от нас облаками. У всех планет-гигантов есть спутники. У Юпитера известно 12 спутников, у Сатурна — 9, у Урана — 5 и у Нептуна — 2. Четыре главных спутника Юпитера были открыты Галилеем. Они хорошо видны с Земли.

APPENDIX E.

```
####################################################################################################
                                  *** INPUT  RUSSIAN  TEXT ***
####################################################################################################

                          #PLANETAXO<SOLN*SISTEMAOG
                          -------------------------

                                  *** PARA NO : 1

#SOLN*SISTEMAOL DEVJATJOO PLANETAXO:#MERKURIJOO VENERAOO ZEMLJAOO MARSOO JUPITEROO SATURNOO URANOO NEPTUNOO PLUTONOO.
PLANETAXO ODELIAT6* DVAOO GRUPPAXK.#MERKURIJOO VENERAOO MARSOO ZEMN*TIPOO PLANETAXO.
#JUPITEROO SATURNOO URANOO NEPTUNOO GIGANTYOO PLANETAXO.PLUTONOO OOTKRYLO* TOLJKOOO<1930XL.
ON2OO ESHCHJOOO OCHENJOO OMALO*IZUCHIT3*+.

                                  *** PARA NO : 2

MERKURIJOO ONAHODI2T3 BLIZHEOO<SOLNCEOD CHEMOO DRUG*PLANETAXA.#MERKURIJOO NEMNOGOOO BOLJSHEOO CHEMOO LUNAOA.
#DIAMETROO<MERKURIJOG 484000 KMOO.#AOO DIAMETROO<LUNAOG 347600 KMOO.#ATMOSFERAXO MERKURIJOL NETOO.

                                  *** PARA NO : 3

RAZMERAOO<MASSAOO<PLOTNOSTJOO<VENERAOG @BLIZKJUT2 ZEMN*TIPOO PLANETAXA.VENERAOO OOKRUZHIS&+ ATMOSFERAOI.
ATMOSFERAOO ODELIT3* OBLAKOXI DVAOO SLOJXK.VERHN*SLOJOL ONAJTIS& BOLJSH*KOLICHESTVOOO UGLEKISL*GAZOO.
ATMOSFERAOO<VENERAOG OOTKRYS& 17610L LOMONOSOVOI.VENERAOO @MEDLENNO*VRASHCHA2ET3 VOKRUGOO<SVOJ*OSJOA.
#VENERAOL DENJOO DLINNEEOO CHEMOO ZEMLJAOA.

                                  *** PARA NO : 4

MARSOO OIZUCHIS&+ BOLJSHEOO LJUDIOI CHEMOO DRUG*PLANETAXA.POVERHNOSTJOO<MARSOG @POSTOJANNO*NABLUDAET3*+ UCHONIJXI.

                                  *** PARA NO : 5

ZEMLJAOR OHOROSHO*VIDIAT6* BEL*SHAPKAXO<POLJUSXG.ZEMLJAOR OHOROSHO*VIDIAT6* MOREXO TIPOO TJOMN*OBLASTJXO.
ZEMLJAOR OHOROSHO*VIDIAT6* LINIJAXO<KANALXG.ZIMAOL CVETOO<POVERHNOSTJON<PLANETAOG @IZMENIA2ET3.
MARSOO @NAHODI2T3 SOLNCEOR DALJSHEOO CHEMOO ZEMLJAOA.#MARSOL KLIMATOO HOLODNEEOO CHEMOO ZEMLJAOA.

                                  *** PARA NO : 6

ATMOSFERAXO OOBNARUZHIP& GIGANTYOO PLANETAXL.ONI2OO OSOSTOIAT6 METANAOK AMMIAKOK VODORODOK GELIJOK.
GLUBINAOO<GAZOV*SLOJOG GIGANTY*PLANETAXL OCHENJOO @VELIKET3.
VNUTREN*SLOJXO<ATMOSFERAOG<JUPITEROL @ISPYTYVAJUT2 OGROMN*DAVLENIEOO<VERHN*SLOJXR<GAZXN.
OCHENJOO<NIZK*TEMPERATURAOO<ATMOSFERAXG GIGANTY*PLANETAXL @POKAZYVAET3 CHTO1OO GIGANTY*PLANETAXO @POLUCHAJUT2 TEPLOOO<SOLNCEOR.
GIGANTY*PLANETAXO ONAHODI2AT6 SOLNCEOR OCHENJOO DALEKOOO.

                                  *** PARA NO : 7

POVERHNOSTJOO<PLANETAXG OZAKRYS&+ OBLAKOXI.VSEOO GIGANTY*PLANETAXL OESAT6 SPUTNIKXO.JUPITEROL OESAT6 1200 SPUTNIKXO.
SATURNOL OESAT6 900 SPUTNIKXO.URANOL OESAT6 500 SPUTNIKXO.NEPTUNOL OESAT6 200 SPUTNIKXO.
CHETYREOO<GLAVN*SPUTNIKXO<JUPITEROG OOTKRYP& GALILEJOI.ONI2OO OHOROSHO*VIDIAT6* ZEMLJAOR.


####################################################################################################
```

APPENDIX F.

##

*** PRE-EDITED RUSSIAN TEXT BY THE PRE-PROCESSOR ***

##

#SOLN*SISTEMAOG PLANETAXO

****

*** PARA NO . 1

#SOLN*SISTEMAOL DEVJATJOO PLANETAXO : #MERKURIJOO VENERAOO ZEMLJAOO MARSOO JUPITEROO SATURNOO URANOO NEPTUNOO PLUTONOO . PLANETAXO ODELIAT6* DVAOO GRUPPAXK . #MERKURIJOO VENERAOO MARSOO ZEMN*TIPOO PLANETAXO . #JUPITEROO SATURNOO URANOO NEPTUNOO GIGANTYOO PLANETAXO . PLUTONOO OOTKRYLO* 1930XL TOLJKOOO . ON200 OMALO*IZUCHIT3*+ ESHCHJOOO OCHENJOO .

*** PARA NO . 2

MERKURIJOO ONAHODI2T3 DRUG*PLANETAXA CHEMOO SOLNCEOD BLIZHEOO . #MERKURIJOO LUNAOA CHEMOO NEMNOGOOO BOLJSHEOO . #MERKURIJOG DIAMETROO 484000 KMOO . #AOO LUNAOG DIAMETROO 347600 KMOO . #ATMOSFERAXO MERKURIJOL NETOO .

*** PARA NO . 3

VENERAOG @BLIZKJUT2 PLOTNOSTJOO RAZMERAOO MASSAOO ZEMN*TIPOO PLANETAXA . VENERAOO OOKRUZHIS&+ ATMOSFERAOI . ATMOSFERAOO ODELIT3* OBLAKOXI DVAOO SLOJXK . VERHN*SLOJOL ONAJTIS& BOLJSH*KOLICHESTVOOO UGLEKISL*GAZOO . VENERAOG OOTKRYS& ATMOSFERAOO 17610L LOMONOSOVOI . VENERAOO @MEDLENNO*VRASHCHA2ET3 SVOJ*OSJOA VOKRUGOO . #VENERAOL ZEMLJAOA CHEMOO DENJOO DLINNEEOO .

*** PARA NO . 4

MARSOO OIZUCHIS&+ DRUG*PLANETAXA CHEMOO BOLJSHEOO LJUDIOI . MARSOG @POSTOJANNO*NABLUDAET3*+ POVERHNOSTJOO UCHONIJXI .

*** PARA NO . 5

ZEMLJAOR OHOROSHO*VIDIAT6* POLJUSXG BEL*SHAPKAXO . ZEMLJAOR OHOROSHO*VIDIAT6* MOREXO TIPOO TJOMN*OBLASTJXO . ZEMLJAOR OHOROSHO*VIDIAT6* KANALXG LINIJAXO . ZIMAOL @IZMENIA2ET3 PLANETAOG POVERHNOSTJON CVETOO . MARSOO @NAHODI2T3 ZEMLJAOA CHEMOO SOLNCEOR DALJSHEOO . #MARSOL ZEMLJAOA CHEMOO KLIMATOO HOLODNEEOO .

*** PARA NO . 6

ATMOSFERAXO OOBNARUZHIP& GIGANTYOO PLANETAXL . ONI200 OSOSTOIAT6 METANAOK AMMIAKOK VODORODOK GELIJOK . GAZOV*SLOJOG @VELIKET3 GLUBINAOO GIGANTY*PLANETAXL OCHENJOO . JUPITEROL @ISPYTYVAJUT2 ATMOSFERAOG VNUTREN*SLOJXO GAZXN VERHN*SLOJXR OGROMN*DAVLENIEOO . GIGANTY*PLANETAXO @POLUCHAJUT2 SOLNCEOR TEPLOOO . CHTO100 @POKAZYVAET3 ATMOSFERAXG NIZK*TEMPERATURAOO OCHENJOO GIGANTY*PLANETAXL . GIGANTY*PLANETAXO ONAHODI2AT6 SOLNCEOR OCHENJOO DALEKOOO .

*** PARA NO . 7

PLANETAXG OZAKRYS&+ POVERHNOSTJOO OBLAKOXI . VSEOO OESAT6 GIGANTY*PLANETAXL SPUTNIKXO . JUPITEROL OESAT6 1200 SPUTNIKXO . SATURNOL OESAT6 900 SPUTNIKXO . URANOL OESAT6 500 SPUTNIKXO . NEPTUNOL OESAT6 200 SPUTNIKXO . JUPITEROG OOTKRYP& GLAVN*SPUTNIKXO CHETYREOO GALILEJOI . ONI200 OHOROSHO*VIDIAT6* ZEMLJAOR .

##

APPENDIX G.

##

**** TRANSLATED VERSION OF THE TEXT IN TAMIL *****

##

CUURIYA MANTALATTUTAIYA KOOLKAL

**** PARA NO. 1

CUURIYA MANTALATTIL ONPATU KOOLKAL : PUTAN VELLI PUUMI CEVVAAY VIYAALAN CANI UURANAS NEPTUUN PULUUTOON.
KOOLKAL IRANTU VAKAIKALAAKA PIRIKKAPPATUKINRANA. PUTAN VELLI CEVVAAY PUUMI POONRA KOOLKAL.
VIYAALAN CANI UURANAS NEPTUUN MIKAPPERIYA KOOLKAL. PULUUTOON 1930KALIL TAAN KANTUPITIKKAPPATTATU.
ITU INNUM MIKA KURAIVAAKA AARAAYAPPATUKIRATU.

**** PARA NO. 2

PUTAN MARRA KOOLKALAI KAATTILUM CUURIYANUKKU ARUKIL IRUKIRATU. PUTAN CANTIRANAI KAATTILUM KONCAM PERITUM.
PUTANUTAIYA VITTAM 4840 KI.MII . AANAAL CANTIRANUTAIYA VITTAM 3476 KI.MII. KAARRUMANTALAMKAL PUTANIL ILLAI.

**** PARA NO. 3

VELLIUTAIYA ATARVU, PARIMAANAM, ETAI PUUMI POONRA KOOLKALAI OTTIRUKINRANA. VELLI KAARRUMANTALATTAAL CUULAPPATULLATU.
KAARRUMANTALAM MEEKAMKALAAL IRANTU ATUKKUKALAAKA PIRIKKAPPATUKIRATU.
VELI ATUKKUIL ATIKA ALAVU KARIYAMILA VAAYU KANTUPITIKKAPPATULLATU.
VELLIUTAIYA KAARRUMANTALAM 1761IL LAMANOSOVAAL KANTUPITIKKAPPATULLATU. VELLI ACCUAI CURRI METUVAAKA CULALKIRATU.
VELLIIL PUUMIAI KAATTILUM PAKAL ATIKANIILAM.

**** PARA NO. 4

CEVVAAY MARRA KOOLKALAI KAATTILUM PERITUM MANITARKALAAL AARAAYAPPATULLATU.
CEVVAAYUTAIYA MEELPARAPPU VINGAANIKALAAL TOTARNTU URRUNOOKKAPPATUKIRATU.

**** PARA NO. 5

PUUMIILIRUNTU TURUVAMKALUTAIYA VELLAI MUKALKAL NANRAAKA PAARKKAPPATUKINRANA.
PUUMIILIRUNTU KATALKAL POONRA IRUNTA PIRATEECAMKAL NANRAAKA PAARKKAPPATUKINRANA.
PUUMIILIRUNTU KAALVAAYKALUTAIYA VARIKAL NANRAAKA PAARKKAPPATUKINRANA.
KULIRKAALATTIL KOOLUTAIYA MEELPARAPPUIN NIRAM MAARUKIRATU.
CEVVAAY PUUMIAI KAATTILUM CUURIYANILIRUNTU TUURATTIL IRUKIRATU. CEVVAAYIL PUUMIAI KAATTILUM TATPAVEPPAM ATIKAKULIR.

**** PARA NO. 6

KAARRUMANTALAMKAL MIKAPPERIYA KOOLKALIL KANTUPITIKKAPPATULLANA.
IVAI METANAALAAKA AMMOONIYAAAAKA NIIRVAAYUAAKA HIILIYATTAAKA IRUKINRANA.
VAAYU ATUKKUUTAIYA AALAM PERIYA KOOLKALIL MIKA ATIKAMAAYIRUKIRATU.
VIYAALANIL KAARRUMANTALATTUTAIYA UL ATUKKUKAL VAAYUKALIN VELI ATUKKUKALILIRUNTU ATIKA ALUTTAM PERUKINRANA.
PERIYA KOOLKAL CUURIYANILIRUNTU VEPPAM PERUKINRANA,
ENRU KAARRUMANTALAMKALUTAIYA KURAINTA VEPPANILAI MIKA PERIYA KOOLKALIL KAATTUKIRATU.
PERIYA KOOLKAL CUURIYANILIRUNTU MIKA TUURATTIL IRUKINRANA.

**** PARA NO. 7

KOOLKALUTAIYA MEELPARAPPU MEEKAMKALAAL MOOTAPPATULLATU. ELLAA PERIYA KOOLKALIL TUNAIKKOOLKAL IRUKINRANA.
VIYAALANIL 12 TUNAIKKOOLKAL IRUKINRANA. CANIIL 9 TUNAIKKOOLKAL IRUKINRANA.
UURANASIL 5 TUNAIKKOOLKAL IRUKINRANA. NEPTUUNIL 2 TUNAIKKOOLKAL IRUKINRANA.
VIYAALANUTAIYA MUKKIYA TUNAIKKOOLKAL NAANKU KALILIYOOAAL KANTUPITIKKAPPATULLANA.
IVAI PUUMIILIRUNTU NANRAAKA PAARKKAPPATUKINRANA.

'+' Represents Source Text Corrected for Spelling While Translating

##

APPENDIX H,

தானியங்கி மொழிபெயர்ப்பு உரை உருசியன் — தமிழ்

சூரிய மண்டலத்துடைய கோள்கள்

சூரிய மண்டலத்தில் ஒன்பது கோள்கள்: புதன், வெள்ளி, பூமி, செவ்வாய், வியாழன், சனி, யூரானஸ், நெப்டியூன், புளூட்டோன். கோள்கள் இரண்டு வகைகளாக பிரிக்கப்படுகின்றன. புதன், வெள்ளி, செவ்வாய், பூமி போன்ற கோள்கள், வியாழன், சனி, யூரானஸ், நெப்டியூன் மிகப் பெரிய கோள்கள். புளூட்டோன் 1930களில் தான் கண்டுபிடிக்கப்பட்டது. இது இன்னும் மிக குறைவாக ஆராயப் படுகிறது.

புதன் மற்ற கோள்களைக் காட்டிலும் சூரியனுக்கு அருகில் இருக்கிறது. புதன் சந்திரனைக் காட்டிலும் கொஞ்சம் பெரிதும். புதனுடைய விட்டம் 4840 கி.மீ. ஆனால் சந்திரனுடைய விட்டம் 3476 கி.மீ. காற்று மண்டலங்கள் புதனில் இல்லை.

வெள்ளியுடைய அடர்வு, பரிமாணம், எடை பூமி போன்ற கோள்களை ஒத்திருக்கின்றன. வெள்ளி காற்று மண்டலத்தால் சூழப் பட்டுள்ளது. காற்று மண்டலம் மேகங்களால் இரண்டு அடுக்குகளாக பிரிக்கப்படுகிறது. வெளி அடுக்கில் அதிக அளவு கரியமில வாயு கண்டுபிடிக்கப்பட்டுள்ளது. வெள்ளியுடைய காற்று மண்டலம் 1761 இல் லமனோசோவால் கண்டுபிடிக்கப்பட்டுள்ளது. வெள்ளி அச்சுவை சுற்றி மெதுவாக சுழல்கிறது. வெள்ளியில் பூமியைக் காட்டிலும் பகல் அதிக நீளம்.

செவ்வாய் மற்ற கோள்களை காட்டிலும் பெரிதும் மனிதர் களால் ஆராயப்பட்டுள்ளது. செவ்வாயுடைய மேல்பரப்பு விஞ் ஞானிகளால் தொடர்ந்து உற்று நோக்கப்படுகிறது.

பூமியிலிருந்து துருவங்களுடைய வெள்ளை முகல்கள் நன்றாகப் பார்க்கப்படுகின்றன. பூமியிலிருந்து கடல்கள் போன்ற இருண்ட பிரதேசங்கள் நன்றாகப் பார்க்கப்படுகின்றன. பூமியிலிருந்து கால் வாய்களுடைய வரிகள் நன்றாகப் பார்க்கப்படுகின்றன. குளிர் காலத்தில் கோளுடைய மேல்பரப்பின் நிறம் மாறுகிறது. செவ்வாய் பூமியை காட்டிலும் சூரியனிலிருந்து தூரத்தில் இருக்கிறது. செவ்வாயில் பூமியை காட்டிலும் தட்ப வெப்பம் அதிக குளிர்.

காற்று மண்டலங்கள் மிகப்பெரிய கோள்களில் கண்டு பிடிக்கப்பட்டுள்ளன. இவை மெதனாலாக, அம்மோனியாவாக நீர் வாயுவாக ஹீலியாதாக இருக்கின்றன. வாயு அடுக்குடைய ஆழம் பெரிய கோள்களில் மிக அதிகமாயிருக்கிறது. வியாழனில் காற்று மண்டலத்துடைய உள் அடுக்குகள் வாயுகளின் வெளி அடுக்குகளி லிருந்து அதிக அழுத்தம் பெறுகின்றன. பெரிய கோள்கள் சூரியனி லிருந்து வெப்பம் பெறுகின்றன, என்று காற்றுமண்டலங்களுடைய குறைந்த வெப்பநிலை மிக பெரிய கோள்களில் காட்டுகிறது. பெரிய கோள்கள் சூரியனிலிருந்து மிக தூரத்தில் இருக்கின்றன.

கோள்களுடைய மேல்பரப்பு மேகங்களால் மூடப்பட்டுள்ளது. எல்லா பெரிய கோள்களில் துணைக்கோள்கள் இருக்கின்றன. வியாழனில் 12 துணைக்கோள்கள் இருக்கின்றன. சனியில் 9 துணைக்கோள்கள் இருக்கின்றன. யூரானசில் 5 துணைக் கோள்கள் இருக்கின்றன. நெப்டியூனில் 2 துணைக்கோள்கள் இருக்கின்றன. வியாழனுடைய முக்கிய துணைக்கோள்கள் நான்கு கலிலியோவால் கண்டுபிடிக்கப்பட்டுள்ளன. இவை பூமியிலிருந்து நன்றாக பார்க்கப் படுகின்றன.

BIBLIOGRAPHY

Andronov M.S (1965) 'Concise Russian Grammar' included in the Russian-Tamil dictionary. Moscow: Soviet Encyclopaedia.

Boitet ch and Nedobejkine Russian-French at Geta:Out line of the method and detailed example. Grenoble:University of Grenoble. F-38041 Grenoble-Cedex 53.

Bowles E.A (1967) Computers in Humanitics Research. U.S.A (New Jersey) Prentice-Hall,Inc.,

Ducrot J.M. 'Titus-4' in Systeme De Traduction Automatique Et Simultanee En 4 Langues. Institut Textile De France.

Ferenc Papp and Gyorgy Szepe (1976)Papers in computational linguistics. Hungary: Mouton.

Ganeshsundaram P.C. Machine Translation of south Asian languages. Bangalore: Indian Institue of Science.

Jacoby L.S. and Kowalik S. (1980)Mathematical Modelling With Computers. U.S.A. Prentice-Hall, Inc.,

Joseph Raben (1977) Computer-Assisted Research in the Humanities. New York: Pergamon Press.

Mathias J. Computer Support to Translation: Foreign Broadcast Information Service. (FBIS).

Namperumal S. (1981) Application of programming Languages for processing natural languages with special reference to English and Tamil. Bangalore: Foreign Language Section. Indian Institute of Science.

Naughton MC Robert (1982) Elementary computability, Format Languages, and Automata. U.S.A. (New Jersey): Prentice-Hall, Inc.,

Pulkina I and Zakhava-Nekrasova E. Russian. Moscow: 'Russky Yazyk' Publishing House.

Sedelow S.Y. and Sedelow A.Jr (1972) Language research and the computer. Kansas: National Technical Information Service.

Smirnitsky A. I (1970) Essentials of Russian Grammar. Moscow: Higher School Publishing House.

U.S.Department of Commerce (1977) Machine-Assisted Translation In West Germany. U.S: National Technical Information Service.

U.S. Department of Commerce (1981) Published Search. U.S: National Technical Information Service (NTIS).